

Eötvös Loránd Tudományegyetem  
Bölcsészettudományi Kar

## DOKTORI DISSZERTÁCIÓ

Deme Andrea

### Az énekelt magánhangzók fonetikai elemzése

Nyelvtudományi Doktori Iskola

vezető: Prof. Dr. Bárdosi Vilmos CSc

Alkalmazott Nyelvészeti Doktori Program

vezető: Prof. Dr. Gósy Mária DSc

A bíráló bizottság tagjai és tudományos fokozatuk:

A bizottság elnöke: Prof. Dr. Gósy Mária DSc

Hivatalosan felkért bírálók: Prof. Dr. Vicsi Klára DSc

Dr. habil. Bóna Judit PhD

A bizottság titkára: Dr. Gyarmathy Dorottya PhD

A bizottság további tagjai: Dr. Szaszák György PhD;

Prof. Dr. Olaszy Gábor DSc, Prof. Dr. Adamikné Jászó Anna DSc (póttagok)

Témavezető: Dr. habil. Markó Alexandra PhD

Társtémavezető: Prof. emer. dr. Johan Sundberg PhD

2015

**ADATLAP**  
**a doktori értekezés nyilvánosságra hozatalához**

**I. A doktori értekezés adatai**

A szerző neve: Deme Andrea

MTMT-azonosító: 10024606

A doktori értekezés címe és alcíme: Az énekelt magánhangzók fonetikai elemzése

DOI-azonosító: 10.15476/ELTE.2015.096

A doktori iskola neve: Nyelvtudományi Doktori Iskola

A doktori iskolán belüli doktori program neve: Alkalmazott Nyelvészeti Doktori Program

A témavezető neve és tudományos fokozata: Dr. habil. Markó Alexandra PhD (egy. docens)

A témavezető munkahelye: ELTE Bölcsészettudományi Kar, Fonetikai Tanszék

**II. Nyilatkozatok**

**1. A doktori értekezés szerzőjeként**

a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom az ELTE BTK Doktori és Tudományszervezési Hivatal ügyintézőjét, Manhercz Mónikát, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.

b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (dátum)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követően egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.

**2. A doktori értekezés szerzőjeként kijelentem, hogy**

a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;

b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.

**3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének Plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.**

Kelt: Budapest, 2015. május 28.

a doktori értekezés szerzőjének aláírása

## Köszönetnyilvánítás

Az értekezés létrejöttében nyújtott segítségéért számos tanáromnak, kollégámnak, barátomnak tartozom hálával.

Köszönöm Gósy Máriának, hogy a doktori tanulmányok irányába terelt, hogy emberileg, szakmailag mindvégig támogatott. Az ő jótékony noszogatása nélkül sohasem adom a fejem tudományos pályára. Köszönöm Surányi Balázsnak, hogy lehetővé tette számomra, hogy elkezdhessem kutatói tevékenységemet az MTA Nyelvtudományi Intézetében (a Lendület projektum keretében). Köszönöm ezen kívül azt is, hogy a támogatásával eljuthattam Svédországba, ahol rengeteg nélkülözhetetlen impulzus ért, és sok fontos szakmai kapcsolatra tehettem szert. Köszönöm Bárkányi Zsuzsannának, Neuberger Tildának, Bóna Juditnak és Vicsi Klárának az értekezéssel kapcsolatos tanácsaikat, kritikai észrevételeiket. Köszönöm Mády Katalinnak, hogy részt vehettem statisztika kurzusain, és köszönöm a sok segítséget az adatelemzésben, a szakmai beszélgetéseket, a tanácsokat, amelyekkel mind hozzájárult szakmai fejlődésemhez. Köszönöm Grácsi Teklának a rengeteg rám fordított időt, köszönöm, hogy tanulhattam tőle. Köszönöm az éjszakába nyúló konzultációkat, az együtt gondolkozást és a támogató segítséget annak a házi dolgozatnak a megírásában, mely az értekezés alapját képezi. Köszönöm Vincze Miklósnak, aki nagyon sok mindent segített megértenem a fizika rejtelmeiből. I acknowledge all the invaluable opportunities I got from Sten Ternström and Svante Granqvist to improve myself. My special recognition goes out to Johan Sundberg for all the support and friendship that I had never dreamed to have. Köszönöm Markó Alexandrának, hogy a szakdolgozatom születésétől fogva témavezetőként támogat. Köszönöm a rengeteg szakmai és baráti segítséget, a lehetőségeket, és főként – a bizalmát. Ő volt az, aki akkor is hitt bennem, mikor én képtelen voltam hinni magamban, és aki nélkül nem csak a dolgozat nem születhetett volna meg, de én sem lennék ma az, aki vagyok. Köszönet illeti a barátaimat, és mindenkit, akivel az elmúlt tíz évben találkoztam. És köszönettel tartozom Káldi Tamásnak is számos dologért, de legfőképpen azért, hogy mindig és mindenben támogatott.

# Tartalom

Előszó .....	8
1. Bevezetés .....	11
1.1. A magánhangzók produkciója és percepciója .....	11
1.1.1. A beszédkeltés folyamata, különös tekintettel a magánhangzók produkciójára.....	11
1.1.1.1. A zöngé (azaz a forrás) keletkezése, az alapfrekvencia.....	12
1.1.1.2. A toldalékcso mint akusztikus szűrő, a formánsok.....	18
1.1.1.3. A magánhangzók artikulációs és akusztikai jellemzése .....	27
1.1.1.4. A magyar magánhangzók artikulációs és akusztikai jellemzése .....	33
1.1.2. A magánhangzók percepciója .....	39
1.1.2.1. Beszédpercepció elméletek .....	41
1.1.2.2. A magánhangzó-azonosítás kísérletes eredményei és a magánhangzó-szerű hangok feldolgozása a percepció rendszerben.....	44
1.1.2.3. A magyar magánhangzók azonosításának kísérletes eredményei.....	52
1.2. Az éneklés és az énekelt magánhangzók fonetikai jellemzése.....	56
1.2.1. Az énekelt magánhangzók artikulációs vizsgálata .....	61
1.2.1.1. A vertikális és horizontális gégehelyzet az éneklésben .....	61
1.2.1.2. A gége belső mozgásai: a nyelvtöcsont szerepe az éneklésben .....	67
1.2.2. Az énekelt magánhangzók akusztikai vizsgálata, rezonanciák a szoprán éneklésben.....	69
1.2.3. Az énekelt magánhangzók percepciója .....	82
1.2.3.1. Az énekelt magánhangzók észlelésével kapcsolatos gyakori feltételezések .....	83

1.2.3.2. Az énekelt magánhangzók észlelésének vizsgálata .....	84
1.3. A jelen kutatás célja, kérdések, hipotézisek .....	107
2. Kísérletek .....	110
2.1. A magyar magánhangzók ejtése és észlelése az éneklésben .....	110
2.1.1. Anyag, módszer és kísérleti személyek .....	110
2.1.2. Eredmények.....	112
2.1.2.1. Észlelési vizsgálat.....	112
2.1.2.2. Akusztikai elemzés .....	118
2.1.3. Következtetések.....	123
2.2. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: a mássalhangzó képzésmódjának hatása .....	127
2.2.1. Anyag, módszer és kísérleti személyek .....	130
2.2.2. Eredmények.....	131
2.2.3. Következtetések.....	136
2.3. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: azonosítás mássalhangzó-környezetben, izolált ejtésben, valamint a zöngéindítás mint akusztikai kulcs hiánya esetén .....	140
2.3.1. Anyag, módszer és kísérleti személyek .....	145
2.3.2. Eredmények.....	147
2.3.3. Következtetések.....	154
2.4. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: a magánhangzók azonosítása értelmes és értelmetlen szavakban.....	158
2.4.1. Anyag, módszer, kísérleti személyek .....	162

2.4.2. Eredmények.....	165
2.4.2.1. Bevezető gondolatok a résztvevők benyomásai alapján.....	165
2.4.2.2. Az ejtési szándéknak megfelelő magánhangzó-azonosítás az értelmes és az értelmetlen szavakban.....	166
2.4.2.3. A magánhangzók az ejtési szándéknak megfelelő azonosítása a két kondícióban .....	168
2.4.2.4. A tévesztések tendenciái az egyes magánhangzók esetében az alaphékvencia függvényében .....	170
2.4.3. Következtetések.....	179
2.5. Az énekelt magánhangzók akusztikai szerkezete néhány módszertani újítás tükrében, valamint az akusztikai és percepcióss jellemzők összefüggései .....	184
2.5.1. Az elektroglottográfia és az elektroglottogram .....	186
2.5.2. Az inverz szűrés, az áramlási glottogram, valamint az elektroglottográfia szerepe az inverz szűrésben .....	189
2.5.3. Traunmüller modellje és a modell predikciói az énekelt magánhangzók azonosítására .....	191
2.5.4. Anyag, módszer és kísérleti személyek .....	197
2.5.4.1. Felvételtészítés és akusztikai elemzések .....	197
2.5.4.2. Percepcióss vizsgálat.....	203
2.5.5. Eredmények.....	204
2.5.5.1. Az énekelt magánhangzók első formánsának alakulása.....	205
2.5.5.2. Az énekelt magánhangzók második formánsának alakulása .....	208
2.5.5.3. Az akusztikai magánhangzótér az alaphékvencia függvényében .....	208
2.5.5.4. A percepcióss adatok és a percepcióss adatok összefüggései az akusztikai eredményekkel.....	210

2.5.6. Következtetések.....	225
3.Összegzés és következtetések .....	238
3.1. A hipotézisek vizsgálata.....	238
3.2. A vizsgálatok közvetlen hozadéka, kitekintés .....	245
3.3. Tézisek.....	248
Irodalom.....	250
A szerző az értekezésben felhasznált publikációinak listája .....	262

## Előszó

A jelen értekezésben bemutatásra kerülő kutatást és kísérletsorozatot egyetlen naiv kérdés indította el, mely az értekezés szerzőjében fogalmazódott meg az éneklés során használt hangmagasságok (alapfrekvenciák) és a magánhangzók akusztikai szerkezetének látszólagos „ellentmondása” nyomán. Az egyes magánhangzók minőségét és a magánhangzók egymástól való elkülöníthetőségét a magánhangzók létrehozásakor az artikulációs szervek állását jellemző üregi rezonanciák, azaz a formánsok alakítják ki. A formánsok között kitüntetett jelentőséggel az első és második formáns bír, ezek a formánsfrekvenciák ugyanis az egyes magánhangzókra jellemzőek, ráadásul az artikulációs szervek állásával viszonylag egyszerűnek tekinthető összefüggésben állnak. Az első formáns a nyelv függőleges helyzetével fordított arányban változik (a magasabb nyelvállás tehát alacsonyabb frekvenciaértékeket eredményez az első formánsra, így pl. az /i/ első formánsa alacsonyabb, mint a nála nyíltabb /a:/ első formánsa), míg a második formáns a vízszintes nyelvhelyzet függvényében alakul (a hátul képzett, veláris hangzók, pl. az /u/ második formánsa alacsonyabb, mint az elől képzett, palatális magánhangzók, pl. az /i/ második formánsa, vö. pl. Gósy 2004). Ennek megfelelően a zárt magánhangzók, így például az /i/ vagy az /u/ első formánsa meglehetősen alacsony frekvenciaértékű – annál a frekvenciaértéknél minden bizonnyal alacsonyabb, mint amilyen alapfrekvencián a magas hangosztályok énekesei, elsősorban a szopránok énekelnek, tehát amely alapfrekvenciákon a szopránok igen gyakran képeznek beszédhangokat.

Felmerül tehát a kérdés, mi történik a magánhangzók első (sőt adott esetben második) formánsával a magas alapfrekvenciájú éneklésben, akkor, ha az énekelt alapfrekvencia magasabb, mint a kérdéses formánsok átlagos (beszédben jellemző) frekvenciaértékei. Ha ugyanis ezek a formánsok „eltűnnek” a magánhangzó spektrumából (tehát az énekelt magánhangzókból csak a magasabb sorszámú formánsok rezonálnak), az azt jelenti, hogy az egyes magánhangzóminőségek is megszűnnek a magas alapfrekvenciákon. Ez mégsem valószínű, hiszen az énekese általában jól érthetőnek tűnő, de feltétlenül nagy hangerejű hangot, nagy intenzitású



magánhangzókat képeznek, ami pusztán a felsőbb sorszámú, ráadásul az artikulációs szervekkel nem is igazán befolyásolható frekvenciájú rezonanciákkal elég nehezen elképzelhető.

Mint utóbb kiderült, a kérdés – természetesen – korántsem tekinthető egyszerűnek, ennek megfelelően a nemzetközi tudományos világban is tekintélyes szakirodalma van. Ezen irodalom tanúsága szerint az énekesek a magas alaphfrekvenciájú éneklés során egészen különleges (és a mindennapi beszélők számára ismeretlen) hatékonysággal használják ki artikulációs szerveik és az artikulációs üregrendszer (elsősorban a garat és a szájüreg) rezonanciáit, ám nem csak azért, hogy a magas alaphangok produkcióját megvalósítsák, hanem azért is, hogy a kiadott hang erőssége elegendő legyen ahhoz, hogy az énekhang erősítés (mikrofon és hangfalak) nélkül is jól hallható legyen akár egy teljes zenekaron át. Mivel azonban az ehhez a különleges teljesítményhez használt rezonanciák egyúttal a magánhangzók kialakításáért is felelősek (lennének), és az énekesek az éneklés során mindenképpen valamely nyelv magánhangzó- (és mássalhangzó-) készletéből építkező szövegeket énekelnek, felmerül a kérdés: mi történik az éneklésben ejtett magánhangzókkal akkor, ha azok specifikus rezonanciáit (formánsait) az énekes inkább a szép és nagy hangerejű énekhangképzés szolgálatába állítja az egyes magánhangzók közti különbség megtartása helyett? Az értekezésben erre a kérdésre keressük a választ mind az énekesek, azaz a produkció és az akusztikai megvalósítás, mind pedig a hallgatók, azaz az észlelés, a percepció oldaláról. Reményeink szerint e kérdés körüljárása révén olyan új oldaláról ismerhetjük meg mind az emberi beszédprodukciós, mind pedig a beszédpercepció működéseket, melyek megvilágítására a beszédvizsgálatok révén nem volna módunk.

Az értekezés a beszéd- és énekhang produkciójához szükséges beszédképző szervek és a beszéd képzésében játszott szerepük áttekintésével kezdődik, melyet előbb a magánhangzók, majd az énekelt magánhangzók már eddig feltárt artikulációs, akusztikai és percepció jellemzőinek leírása követ. A értekezés első fejezetének végén, mindennek zárásképp újabb megválaszolásra váró kérdéseket vetünk fel az éneklés produkciós és percepció vetületét illetően, különös tekintettel a magyar magánhangzók megvalósítására az éneklésben. A értekezés második részében a kérdések megválaszolásához lefolytatott kísérleteinket és azok eredményeit ismertetjük. Végezetül, az értekezés befejezéseként a kísérleti eredmények összefoglalásával

az értekezésben bemutatott kutatómunka új eredményeit vesszük sorra, majd zárásként következtéseket vonunk le belőlük.

# 1. Bevezetés

## 1.1. A magánhangzók produkciója és percepciója

### 1.1.1. A beszédkeltés folyamata, különös tekintettel a magánhangzók produkciójára

A beszédkeltés alapvető fiziológiai szervei a tüdő, a légcső, a gége és az abban található hangszalagok, a garat, a szájüreg valamint az orrüreg. A hangszalagoktól (vagyis a gégetől) kezdve a száj- és az orrüregig bezárólag ezek a szervek együttesen alkotják a szupraglottális üregrendszert, melyet átfogóan artikulációs csatornának, toldalékcsőnek vagy ritkábban (a *vocal tract* terminus magyarításával) vokális traktusnak is nevezzük. Ezzel párhuzamosan a gége, potnosabban a hangrés alatti üregrendszer neve szubglottális üregrendszer (Gósy 2004).

A beszédkeltés mechanizmusát Fant (1960) úttörő munkája óta hagyományosan lineáris, egy forrásból és szűrőből álló rendszerben képzeljük el (source–filter theory, vö. pl. Gósy 2004, Vicsi 2010b). Ebben a modellben a hangadás alapja a forrás (valamilyen periodikus vagy nem periodikus, azaz zörejszerű, szabálytalan rezgés), mely a szűrőn átjutva (az ott ért módosítások után) a szájon át kisugárzik – ez adja a hallható beszédet. A linearitás itt arra utal, hogy a beszédképzés folyamatában a hangadás alapját képező levegő, valamint a létrejövő hanghullámok a hangforrás felől a beszédképző szervek rendszerén át csak egy irányba terjednek. Ez bár a valóságban nem pontosan így van, de a modell számára mégis jó közelítést ad az artikulációs csatorna működéséről, valamint elegendő a létrejövő akusztikai jel alapvető jellemzőinek megértéséhez. A forrás és szűrő elnevezések arra utalnak, hogy a hangadást, beszédkeltést valamilyen hangforrás (a magánhangzók esetében ez a zöngé) létrehozásával, majd pedig a létrehozott hang módosításával, szűrésével modellezzük (l. Vicsi 2010b).

A magánhangzók esetében a beszédkeltés során a tüdőből kiáramló levegő megrezgeteti a gégeben található hangszalagokat (ez lesz tehát a forrás), az így megrezgetett levegő (másnéven zöngé) pedig a toldalékcsövön keresztülhaladva különböző akusztikai módosulásokon esik át (a toldalékcső, tehát a szűrő hatására), mígnem végül a szájnyílást elhagyva el-

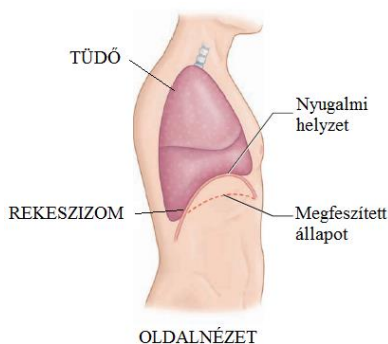
nyeri végleges formáját, és magánhangzóvá alakul. Az itt következőkben ezt a folyamatot, azaz a beszédkeltés közismert fiziológiai folyamatait tekintjük át röviden a forrás és a szűrő, tehát a zönge és a toldalékcső által okozott akusztikai változások elválasztásával, különös tekintettel a magánhangzóképzésre Gósy (2004), Vicsi (2010b), valamint Seikel és munkatársai (2010) alapján.

#### **1.1.1.1. A zönge (azaz a forrás) keletkezése, az alapfrekvencia**

A beszédkeltés a belégzéssel kezdődik, mely során a tüdő megtelik levegővel. A belégzésben a tüdő szerepe passzív, tehát a levegőnek a tüdőbe jutása nem a tüdő öngerjesztett mozgásának eredménye. A tüdő a mellkasban helyezkedik el, két lebernyege a mellkasba a légcsőhöz alulról csatlakozva lóg. A tüdőhöz közvetlenül nem kapcsolódnak izmok, melyekkel mozgatható lenne, közvetve azonban, a kettős falú mellhártya révén mégis szoros összeköttetésben áll a bordákkal, valamint a tüdőt alulról megtámasztó rekeszizmokkal is. A tüdőt a mellkashoz és a rekeszizomhoz is negatív nyomás, azaz egyfajta szívó hatás rögzíti, mely a mellhártya két rétege között áll fenn. A mellhártya egyik rétege a tüdőhöz, míg a másik a mellkashoz kapcsolódik. A szívó hatás ezt a két réteget húzza össze. A mellhártyának köszönhetően tehát mozgatható, pontosabban kifeszíthető a tüdő – ugyanis a mellkas és a rekeszizom mozgását a mellhártya által (passzívan) a tüdő is követi. Mivel a tüdő tehát fentről és oldalról a bordákhoz, alulról pedig a rekeszizomhoz rögzített, azok ellentétes irányú mozgása a tüdő rugalmas lebernyeit kifeszíti, ezzel megnövelve annak térfogatát. Belégzéskor éppen ez történik: a rekeszizom megfeszül, lefelé húzódik, a mellkas pedig a bordaközi izmok megfeszítése révén megemelkedik (1. ábra). Így belégzéskor a tüdő kifeszül, és térfogata megnő. A belégzés, azaz a levegőnek a tüdőbe juttatása ugyanazon az elven alapszik, mint az injekcióstű működése: mivel a gázok mindig igyekeznek kitölteni a rendelkezésükre álló teret, ha a tér kitágul, a benne lévő gáz is tágulni kezd, ami pedig nyomáscsökkenést, azaz egyfajta szívó hatást eredményez. A tüdő a felnőttek esetében nyugalmi állapotban is ki van feszítve a mellkasban, így ebben az állapotában is viszonylag sok levegőt tartalmaz. Gyermekek esetében azonban a tüdő nyugalmi állapotában tapasztalható nagyobb fokú ernyedtsége miatt ez a levegőmennyiség jóval kisebb. Amikor belégzéskor a bordák és a rekeszizom működése révén a tüdő kitágul, az abban

lévő levegő is tágulni, a tüdőben lévő nyomás pedig csökkenni kezd. Az alacsony nyomás kiegyenlítődsre törekszik, és mivel a tüdő a hangszalagok nyitott állásában nem képez teljesen zárt rendszer, így a belégzéskor a fellépő nyomáscsökkenés kiegyenlítődhét. A kiegyenlítőds során a nyílásnál, azaz a hangszalagoknál lévő (a száj- vagy orrnyíláson keresztül a toldalékcsőbe jutott) levegő a tüdőbe áramlik.

A légzés funkcióját tekintve lehet élettani légzés, mely során pusztán az életben maradáshoz szükséges gázcsere zajlik, valamint beszédlégzés is, mely során a belégzést hangadás, beszéd-keltés követi. Előbbi esetben a kilégzés alapvetően passzív folyamat, melyet a megfeszített bordaközi izmok és a rekeszizom izmok elernyesztése, a mellkas lesüllyedése idéz elő. A beszédképzés során azonban a levegő kiengedése fokozatos (és a belégzési szakasznál hosszabb időbeni lefutású), amit a beszélő aktívan is befolyásol a belső bordaközi izmok, valamint a hasfal izmainak munkája révén. A beszédkeltés közbeni aktív kilégzés tehát biztosítja a levegőtartalék folyamatos adagolását.

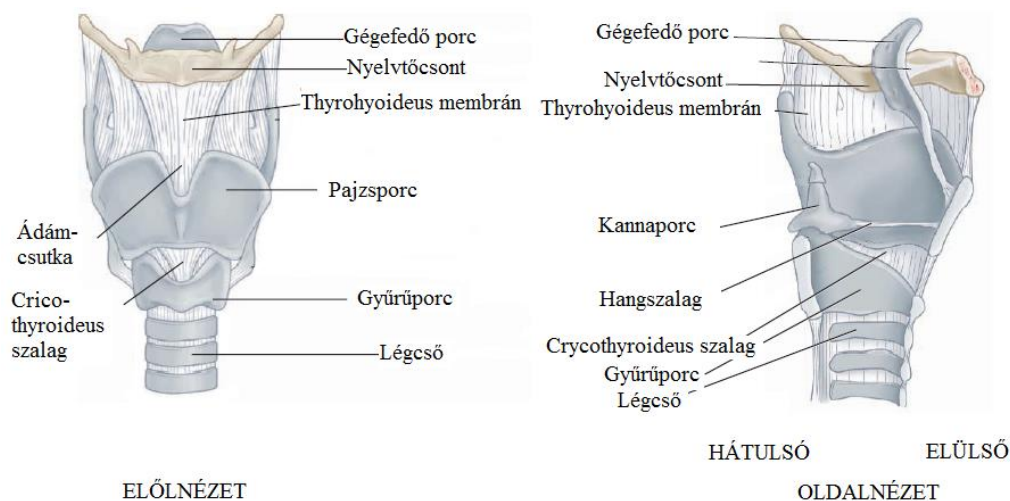


1. ábra: A rekeszizom működése (Seikel et al. 2010: 105 alapján)

A beszédképzés során a belégzés következtében a tüdőben felgyűlt levegő kipréselődik a tüdőből, és a légcsövön keresztül a gégebe jut. A légcső C alakú porcokból álló hajlékony, mégis kemény falú cső, így a nyak szabadon mozgatható, miközben a levegő közel veszteségmentesen, azaz sebességcsökkenés nélkül áramolhat át a légcsövön. A gége a légcső tetején helyezkedik el, porcokból, izmokból, ízületekből és szalagokból áll. Vázát öt porc adja: a gyűrűporc, a pajzsporc, a kannaporcok és a gégefedő porca (l. 2. ábra). Ezeket a porcokat számos

izom és szalag szövi át, így a gége változatos mozgásokra képes. A gégeben húzódnak a hangszalagok, melyek elülső oldala a pajzsporchoz kapcsolódik, hátulsó oldala pedig a kannaporcokhoz ízesül. A hangszalagok előlről mereven rögzítettek, hátulról azonban a kannaporcok mozgása révén változtatható a beállításuk – így a két hangszalag közelíthető (addukció) és távolítható (abdukció).

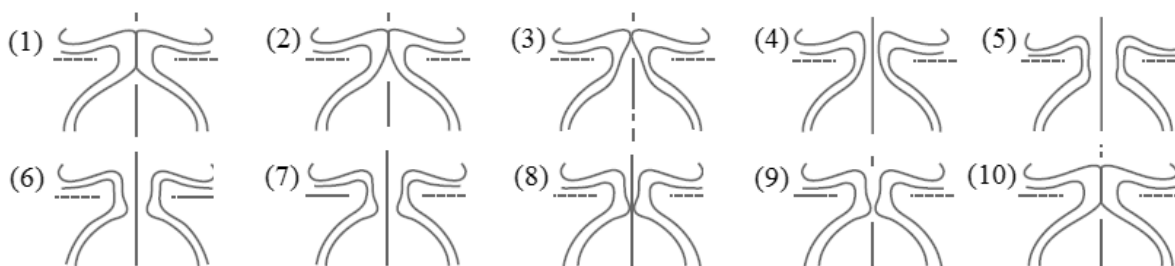
A gége a beszédképzésen kívül számos élettani funkcióval bír, a levegő visszatartása, valamint a hangszalagok közti rés, a hangrés vagy glottisz bezárása szükséges többek közt a mellkas fixálásához (így a mellkas és a karok nagyobb erő kifejtésre képesek), a levegőtartalék visszatartásához (így lehetünk képesek a víz alatt úszni), valamint megakadályozza az idegen anyagoknak a tüdőbe jutását. A gége élettani szempontból elsődleges szerepe a tüdőhöz vezető út védelme és szükség szerinti gyors lezárása – jól mutatja ezt, hogy a hangrés zárásához 3 pár adduktor izom specializálódott, míg a hangrés nyitását mindössze egy pár abduktor izom végzi. A glottiszt védő szereppel bír a felette található gégefedő vagy epiglottisz, egy levél alakú porc, mely nyeléskor a hangszalagrést lezárva megakadályozza a szájtartalom tüdőbe jutását.



2. ábra: A gége felépítése (Seikel et al. 2010: 167 alapján)

A normál módú beszédképzés során a hangrés a zöngés, valamint a zöngétlen hangadáshoz zárt vagy nyitott állásban áll. A hangszalagok lehetséges beállításai közül itt most csak a (magyar) magánhangzók képzésének szempontjából fontos, a zöngképzés során használt zöngé-

állást és annak a zöngéképzésben játszott szerepét emeljük ki. A zöngés hangok, tehát például az összes magyar magánhangzó képzésekor a hangszalagok a hangadás kezdetekor, a zöngeciklus létrehozásának kezdetén zárt pozícióban állnak, azaz a két hangszalag közti rés, a glottisz vagy hangrés zárva van (l. 3. ábra, 1. panel). E miatt az akadály miatt a tüdőből kiáramló levegő nem áramolhat tovább, hanem felgyűlik a zárt hangszalagok alatt. Ha a felgyűlt levegő mennyisége és az így létrejött nyomásnövekedés megfelelő mértékű (illetve a szubglottális nyomás kellő mértékben nagyobb, mint a szupraglottális), a levegő szétfeszíti a hangszalagokat, és tovább áramlik a száj- vagy az ornyílás felé. A hangszalagok távolodása alulról felfelé hullámként terjed (lásd 3. ábra, 2–5. panel). Mivel a hangszalagok a gégében egy szűkületi ponton helyezkednek el, az itt átáramló levegő a légcső és a toldalékcső felsőbb szakaszához képest gyorsabban halad. Ez pedig Bernoulli törvénye értelmében azt vonja magával, hogy amikor a levegő áthalad az éppen csak szétfeszített hangrésen, a hangszalagok közötti nyomás lecsökken, ez pedig szívó hatást fejt ki a légjárat falaira, azaz a hangszalagokra. A szívó hatás következtében a szétfeszítés után a hangszalagok újra összetapadnak, így a zöngéképzés egy új ciklusa kezdődhet meg. A hangszalagok közeledése a Bernoulli-effektus hatására szintén alulról felfelé terjedő állapot (lásd 3. ábra, 6–10. panel). A hangszalagok ciklikus összezáródása és szétnyílása addig tart, amíg elegendő levegő áramlik ki a tüdőből ahhoz, hogy szétfeszítse a zöngéképzéshez, vagyis a fonációhoz összezárt hangszalagokat, majd az így keletkezett résen átáramló levegő ismét össze tudja szívni azokat (kellően nagy sebesség, azaz megfelelően alacsony nyomás és a szub- és szupraglottális terek között fennálló megfelelő nyomáskülönbség segítségével).



3. ábra: A zöngéképzés egy ciklusa (Flemming 2005 alapján)

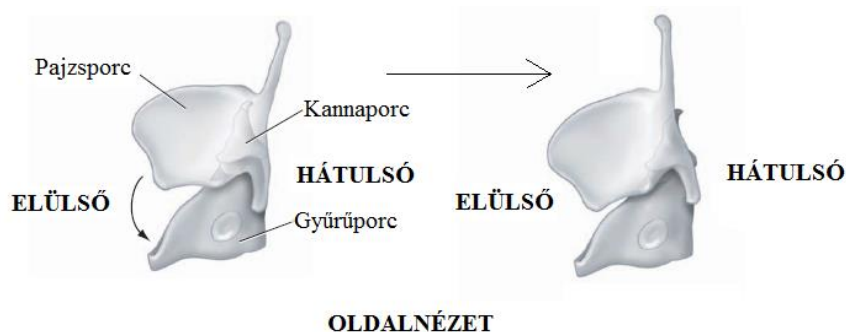
A hangszalagok nyitódása és záródása következtében a tüdőből még egyenletesen áramló levegő kváziperiodikus löketekben hagyja el a géget, így hangnyomásingadást okoz, mely az idő függvényében (a hangszalagok ciklikus összezáródási fázisával összhangban) tehát kváziperiodikus lesz. Az így létrejött nyomásingadozást, azaz hangot nevezzük zöngének, a zöngéképzéshez szükséges rezgést pedig más szóval fonációnak is nevezzük. A zöngé karakterisztikus jellemzője az alaphfrekvenciája, azaz a zöngéképzés során az egy időegység alatt lezajlott ciklusok száma. Ez adja a zöngés magánhangzónak az emberi fül számára is érzékelhető alaphangját, hangmagasságát – míg a rezgés amplitúdója, „nagysága” a hangerősségérzetet befolyásolja. Az alaphfrekvencia jelölése általánosan  $f_0$ , (fizikai, lineáris) mértékegysége Hertz [Hz].

Bár úgy tűnik, az alaphfrekvenciára áttételesen számos tényező hatással van, közvetlenül gyakorlatilag mindegyik a hangszalagok méretét és egységnyi hosszra jutó tömegét befolyásolja, ez tehát az a jellemző, amely az alaphfrekvenciát meghatározza. Az alaphfrekvenciát meghatározó tényező lehet az életkor és a nem (hiszen a hangszalagok hossza és tömege általánosan különbözik a gyermek, a női és a férfi beszélők esetében), a hangszalagok feszítettsége, a gége vertikális és horizontális helyzete (ez ugyanis szintén a hangszalagok feszítettségére hat) vagy az ejtett beszédhang minősége (mely pedig szintén összefüggést mutat a függőleges gégehellyel, így a hangszalagok feszítettségével is).

A hangszalagok mozgását és annak sebességét elsősorban és közvetlenül csak a hangszalagok méretei, valamint az aerodinamikai hatások befolyásolják, mégis, ezekre közvetve az egyes gégeizmok működése is fontos hatással van. Az alaphfrekvencia, azaz az alaphangmagasság emelését ugyanis a hangszalagok megfeszítésével, megnyújtásával érjük el, így tudjuk ugyanis csökkenteni a hangszalagok egységnyi hosszra eső tömegét – míg az alaphfrekvencia csökkenését a hangszalagok ellazítása okozza. A hangmagasság emeléséhez a gyűrűporcot és a pajzsporcot összekötő cricothyroideus izom, valamint a pajzsporcot és a kannaporcokat összekötő belső thyrovocalis izom megfeszítésére van szükség. Az előbbi megfeszítésének hatására a pajzsporc előrebillen, a hangszalagok megnyúlnak, így a zöngéképzés során létrehozott hang alaphfrekvenciája is megemelkedik. Az utóbbi megfeszítése (antagonistaként) pedig szintén ezt a folyamatot segíti.



Az alaphfrekvencia csökkentéséhez az ellenkező irányú manipuláció szükséges, tehát a hangszalagok ellazítása. Ezt a pajzsporcot és a kannaporcokat összekötő külső thyromuscularis izom megfeszítésével érjük el, az ugyanis közelebb húzza egymáshoz a pajzsporcot és a kannaporcokat, így rövidíti, lazítja a hangszalagokat. Mivel a magasabb alaphfrekvencia képzéséhez megfeszített hangszalagokat nehezebb mozgásba (azaz rezgésbe) hozni, az alaphfrekvenciaemeléshez a szubglottális nyomás emelése is szükséges. Ebből következően a magasabb alaphfrekvenciájú hangok ejtését mindig magasabb szubglottális nyomás, azaz magasabb hangnyomásszint, nagyobb hangerősség is kíséri.



4. ábra: A gégeporcok mozgásai az alaphfrekvencia emeléséhez: az alaphfrekvencia emeléséhez a pajzsporc előrebillen, így jobban megfeszíti az előlről a pajzsporchoz, hátulról pedig a kannaporcokhoz csatlakozó hangszalagokat (Seikel et al. 2010: 185 alapján)

A gégében keletkező zöngé tehát meghatározza a keletkező beszédhang alaphfrekvenciáját, azaz annak észlelhető hangmagasságát. Ugyanakkor a zöngé nem csak a hangszalagok rezgésszámára jellemző alaphfrekvenciát, hanem annak egész számú többszöröseit, azaz a felharmonikusokat, felhangokat is tartalmazza.<sup>1</sup> Ezek jelölése többféle lehet, melyek közül a jelen értekezésben az egyszerű átláthatóság kedvéért, valamint azért, hogy a jelölés felidézze a felhangoknak az alaphfrekvenciával való kapcsolatát, az alaphfrekvenciát  $f_0$ , annak felharmonikusait  $2f_0, 3f_0, 4f_0, \dots$  jelöljük.

<sup>1</sup> A felharmonikusok létrejöttének oka egyébként az, hogy a felnyíló glottiszon átáramló levegőmennyiség áramlási sebessége valójában a levegő tehetetlensége miatt nem szinuszos, hanem elferdült jelalakban leírható. Ez az elferdült jelalak viszont már egy komplex hang hullámformája: nem csak a rezgés alaphfrekvenciáját, hanem annak egész számú többszöröseit is tartalmazza (vö. Vicsi 2010).

kusait pedig  $2f_0$ ,  $3f_0$ ,  $4f_0$  stb. notációval jelöljük, a legalacsonyabb alaphfrekvenciájú összetevőtől (az alaphfrekvenciától) a magasabbak felé haladva. A zöngében a felharmonikusok amplitúdója egyenletes, 12 dB/oktáv csökkenéssel van jelen. Ez jellemzi tehát a zöngé spektrumát, azaz a zöngé frekvencia-összetevőinek amplitúdóját a frekvencia függvényében – a folyamat e pontján még elvben (a lineáris forrás-szűrő modell szerint) az ejteni szándékozott magánhangzó minőségétől függetlenül. Ahhoz, hogy a zöngéből különböző magánhangzószerű hangok keletkezzenek, szükséges, hogy a zöngé keresztülhaladjon a toldalécsövön, mely akusztikus szűrőként változtatja meg a zöngének az ember számára hangszíneként érzékelhető spektrumát, azaz a zöngé egyes felhangjainak intenzitását.

#### **1.1.1.2. A toldalécső mint akusztikus szűrő, a formánsok**

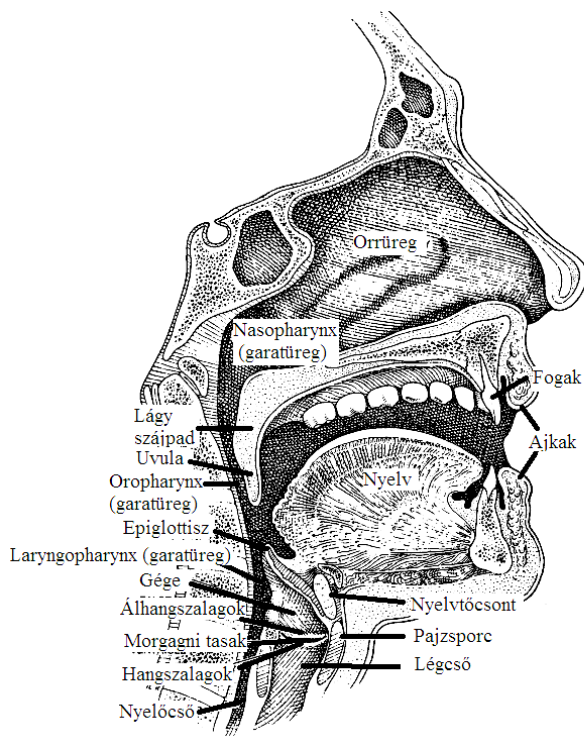
A toldalécső főbb részei a garatüreg (pharynx), a szájüreg (cavum oris) és az orrüreg (cavum nasi). Ezekben az üregekben passzív és aktív artikulációs szervek találhatók, melyek a toldalécső alakját módosítva változtatják annak sajátfrekvenciáit, azaz a toldalécső szűrőtulajdonságait. A jelen fejezetben előbb az artikulációs szerveket és főbb beszédbeli funkcióikat tekintjük át – különös tekintettel a magánhangzók képzésében részt vevő szervekre –, majd az üregi rezonanciák keletkezését tárgyaljuk.

Az aktívan mozgatható artikulációs szervek közül a legnagyobb és talán a legfontosabb a nyelv és az állkapocs, melyek működése lévén a szájüregben kisebb vagy nagyobb szűkületet, illetve rést vagy zárat képzünk a nyelv és a szájpad között a beszédhangok ejtése során. A nyelv mozgásai igen változatosak: a nyelv a külső izmok munkája révén vízszintesen és függőlegesen is mozdítható szerv (a külső izmok határozzák meg tehát a nyelv helyzetét a szájüregben), míg a belső izmok segítségével alakítható a nyelv alak (domborítható, egyenesíthető a nyelvhát stb.). A nyelv mozgása révén befolyásolja a garat méretét, valamint a nyelvet a nyelvtőcsonttal összekötő izmok révén a gége helyzetét is. A nyelv főbb részei a nyelv csúcsa vagy nyelvhegy (apex/cacumen), a nyelvperem vagy nyelvpárkány (corona), nyelvhát (dorsum), valamint a nyelvgyök (radix). A nyelvhátat szokás további három részre is tagolni, ezek az elülső (predorzális), a középső (mediodorzális) és a hátsó (posztdorzális) részek. Bár a nyelv mozgása (elsősorban a függőleges helyzetének módosítása) és az állkapocs nyitásszögé-

nek változtatása szorosan összefügg, az egyik helyzete nem törvényszerűen determinálja a másikat, hiszen bizonyos mértékig külön-külön is mozgatható artikulátoroktól van szó – erre a későbbiekben még a magyar magánhangzórendszer taglalásánál, valamint az éneklés artikulációjának és akusztikai szerkezetének bemutatásánál is utalunk.

Az aktív, mozgatható artikulátorok közé tartoznak még az ajkak, a garatüreg fala, valamint a gyakorlatilag csak izmokból álló lágy szájpad (mely a szájpadlás hátulsó része, a velum palatinum), és annak zárlata a nyelvcsap vagy uvula. Az ajkak legfőbb mozgásai a távolodás vagy a közeledés (egészen a teljes zárképzésig), valamint az ajkak kerekítése és széthúzása, illetve előrecsücsörítése. A lágy szájpad és az uvula legfőbb funkciója, hogy az orrüreg és szájüreg közti átjáró lezárásán vagy megnyitásán keresztül elzárja vagy megnyitja a levegő útját az orrüreg felé, így rendre az orális és nazális hangminőségek létrehozásáért felelős.

A passzív vagy nem mozgatható artikulátorok közé tartoznak a fogak, a fogmeder és a kemény szájpad (mely a szájpadlás elülső része, palatum durum). Ezek azok a szervek, melyekhez az aktív, mobilis artikulátorok közelednek a beszédképzés során (vö. 5. ábra).



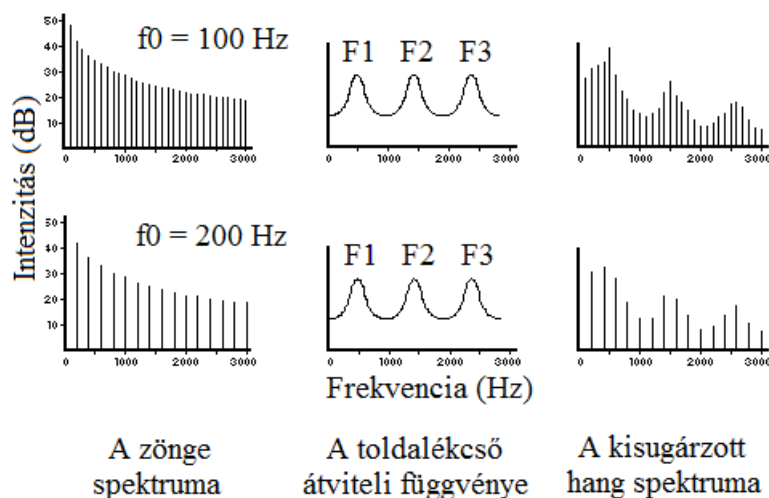
5. ábra: A toldalékcső részei (Sundberg 1977: 83 alapján)

A magánhangzók ejtésekor a nyelv jellemzően nem érintkezik a szájpaddal, csak közelít hozzá, a levegő a toldalékcsőben akadálymentesen áramlik (a rés vagy zár létrehozása elsősorban a mássalhangzók képzésére jellemző). Megjegyzendő mégis, hogy a magánhangzók képzésében is szerepe van annak, hogy a nyelv egyes részei érintkeznek a szájüreg más (immobilis) artikulációs szerveivel, mert ezek a szenzoros visszajelzések artikulációs sarokpontokat jelentve feltehetőleg hozzájárulnak ahhoz, hogy a beszélő az egyes hangok ejtéséhez szükséges artikulációs beállításokat pontosan eltalálja (Stevens 1998). Az orális magánhangzók – így például a magyar magánhangzók – képzése szempontjából a legfontosabb artikulációs szerveknek a szájüregben létrehozott szűkület méretét befolyásoló nyelvet és állkapcsot, valamint az ajkakot tekinthetjük.

A toldalékcső mint üregrendszer más üregekhez hasonlóan rendelkezik bizonyos sajátfrekvenciákkal, melyekkel gerjesztve rezonálni kezd. Ezek a frekvenciák együttesen adják meg a toldalékcső átviteli karakterisztikáját, azaz azt a függvényt, mely a sajátrezonanciákat „összesíti”. Amint azt már említettük, a zöngé mind az alapfrekvenciáján, mind annak egész számú többszörösein tartalmaz frekvenciaösszetevőket, felharmonikusokat. Ennek értelmében tehát a mindennapi (nagyon átlagosan körülbelül 100–400 Hz közti alapfrekvenciájú, vö. Gósy 2004) beszéd során a zöngé olyan akusztikai szerkezetű, hogy a toldalékcsőre jellemző frekvenciahelyeken akár több felharmonikussal is rendelkezik, így gerjeszti azt. Ha tehát van fonáció, és zöngét képezünk, gerjesztjük a toldalékcsővet, mely az alakjának megfelelő frekvenciákon a zöngével együtt rezonálni kezd. Ennek a rezonanciának eredményeképpen pedig a szájnál kisugárzott hang spektruma eltérő lesz annak a hangnak a spektrumától, mint ami a gégénél létrejött – azaz a szájnál kisugárzott hang másféle akusztikai szerkezetű, mint a zöngé. A toldalékcső ugyanis a rezonancia révén megszűri a zöngéhangot: a toldalékcső felerősíti a zöngében a sajátfrekvenciáinak frekvenciasávjába eső felhangokat vagy felhangnyalábokat, és csillapítja az azoktól távolabb eső frekvenciaösszetevőket (azaz csökkenti azok amplitúdóját). Ezt a folyamatot szemlélteti a 6. ábra.

A beszéd szempontjából a toldalékcső rezonanciahelyeinek vagy sajátfrekvenciáinak kiemelt jelentősége van, hiszen ezek határozzák meg a kisugárzott hang specifikus hangszínét, azaz az egyes beszédhangokat. A toldalékcső ezen állóhullámú rezonanciáinak elnevezése a

beszédtudományban formáns, jele pedig F1, F2, F3 stb. a legalacsonyabb frekvenciaértékű rezonanciától kezdve a sorszámozást.



6. ábra: A zöngés magánhangzók létrehozásának folyamata a forrás-szűrő modellben egy mélyebb (felső panel) és egy magasabb (alsó panel) alapfrekvenciájú hang esetében

(<http://www.haskins.yale.edu/featured/heads/mmsp/acoustic.html> alapján)

Mivel a felharmonikusok sűrűsége az alapfrekvenciától függ, ezért elmondható, hogy minél mélyebb a beszédhang, annál több frekvenciakomponense esik a toldalékcső sajátrezonanciáinak frekvenciasávjába, tehát annál több felharmonikusa vesz részt a beszédjel akusztikai kimenetében található frekvenciamaximumok létrehozásában. Ezt szemlélteti a 6. ábra alsó és felső panele: az alsó panelen egy magasabb, 200 Hz alapfrekvenciájú, míg a felső panelen egy alacsonyabb, 100 Hz alapfrekvenciájú hang keletkezését látjuk. A formánsok és a toldalékcső sajátfrekvenciáinak az imént megfogalmazott összefüggése a magas alapfrekvenciájú éneklésben kiemelt jelentőségű (így, mint látni fogjuk, a jelen értekezésben is kiemelt figyelmet kap). A magas alapfrekvencia mint akusztikus tulajdonság ugyanis problematikusá teszi mind a beszédhangok formánsainak detektálását, mind pedig a magas alapfrekvencián ejtett beszédhangok percepciójának értelmezését. Ezekről a későbbiekben még részletesebben is szót ejtünk.

Annak érdekében, hogy a különböző beszédhangok előállítását és a beszédhangokra jellemző frekvenciák kialakulását megérthessük, a beszédkutatók számos modellt alkottak. Ezekben közös, hogy a toldalékcsovet egy vagy több (ez esetben különböző keresztmetszeti területű) csőhöz hasonlítják, ezzel ugyanis jó közelítés adható a toldalékcsőnek az egyes magánhangzók ejtésekor jellemző alakjáról. Így a magánhangzókra jellemző frekvenciahelyeket, azaz formánsokat a valóságban igen bonyolult szerkezetű toldalékcső helyett egy annál jóval egyszerűbb szerkezetű csőből vagy csőrendszerből vezethetjük le. Ezen elméletek a toldalékcsovet általánosan egy az egyik oldalán (a gégénél) zárt, a másik oldalán (a szájnyílásnál) pedig nyitott csőnek tekintik. Az ilyen csőben a gerjesztés hatására ébredő állóhullámú rezonanciák, azaz a cső sajátfrekvenciái adott keresztmetszet mellett csak a cső hosszától függenek. A cső állóhullámú rezonanciái 17 cm hosszúság (azaz egy átlagos férfi toldalékcsőméret) esetén a peremfeltételek figyelembevételével (tehát figyelembe véve, hogy a cső egyik oldalán zárt, míg a másik oldalán nyitott rendszer) megközelítőleg 500, 1500 és 2500 Hz.

A *perturbációs elmélet* a toldalékcsovet egyetlen (például a fenti példában megjelölt sajátrezonanciákkal rendelkező) csőként képzei el, melynek rezonanciáit a beszéd során a beszélő folyamatosan módosítja, mégpedig úgy, hogy a rezonanciák minimum- és maximumhelyeinél (azaz a maximális vagy minimális térfogasebességeknél) szűkíti a toldalékcsovet. Az elmélet szerint ugyanis a térfogatsebesség maximumhelyeinél okozott szűkület csökkenti az adott rezonancia értékét, míg a minimumhelyeknél okozott szűkület növeli azt (Chiba és Kajiyama 1941, idézi Flemming 2005).

A *beszédképzés gerjesztett cső modellje* – vagy röviden az *akusztikus csőmodell* – azonban a magánhangzók ejtésekor létrehozott csőkonfigurációt nem egy, hanem két akusztikusan csatolt csővel modellálja, melyek hossza és keresztmetszete az egyes magánhangzók ejtésekor jellemző toldalékcsőformának megfelelően különböző lehet. E modell szerint az egyes magánhangzók eltérő formánsszerkezetét a két csatolt cső a csőhosszoktól függő sajátrezonanciái, valamint a két cső csatolásának mértéke alakítja ki, ami a két cső keresztmetszeti területeitől, illetve azoknak az arányától függ (Stevens 1998).

Amint azt láttuk tehát, a formánsok frekvenciáit és sáv szélességét a toldalékcső méretei, azaz a toldalékcső hossza és alakja, vagyis az egyes szakaszokon jellemző keresztmetszeti te-

rülete határozza meg. Bár ezek a paraméterek alapvetően minden beszélő esetében fiziológiai-  
lag adottak, mégis bizonyos kereteken belül változtathatók, éppen ezért az üregrendszer saját-  
frekvenciái nem állandók.

A toldalékcső hossza egy adott beszélő esetében egy adott életkorban csak kisebb mérték-  
ben befolyásolható, a változás a függőleges gégehelyzet módosításával (a gége lesüllyesztésé-  
vel vagy felemelésével), illetve az ajkak csücsörítésével érhető el. A toldalékcső hosszát azon-  
ban nagyban megváltoztathatja a testi növekedés.

A toldalékcső keresztmetszeti területe azonban éppen a jelen alfejezet elején taglalt artiku-  
lációs szervek mozgásai révén viszonylag nagymértékben változtatható – a toldalékcső alakjá-  
nak változtatásával tehát valójában annak formánsfrekvenciáit (is) változtajtuk. Ez tehát a tol-  
dalékcső azon tulajdonsága, melyet a beszédképzés során is kihasználunk ahhoz, hogy egy-  
mástól eltérő spektrumú, azaz eltérő hangszínű, minőségű magánhangzókat képezzünk. Az  
egyes magánhangzók ejtése ilyen módon felfogható a rezonátorüreg egy adott artikulációs  
konfigurációra hangolásával, aminek következtében a toldalékcső sajátfrekvenciái adott, az  
egyes magánhangzókra jellemző értékeket vesznek fel. Ennek megfelelően a magánhangzókat  
artikulációsan és (az artikulációs működésekkel szoros összefüggésben) akusztikailag is jelle-  
mezhetjük. Ezt tárgyaljuk a következő alfejezetben.

Mielőtt még azonban áttérnénk a magánhangzók artikulációs és akusztikai jellemzésére,  
fontosnak tartjuk két fogalom rövid tisztázását. Az egyik a hangszín és a hangszínezet elkülö-  
nítése, a másik pedig a formáns fogalmának tisztázása.

A hangoknak az ember számára érzékelhető minőségét *hangszínn*ek hívjuk. A hangszín  
pszichoakusztikai fogalom, az ANSI (American National Standards Institute, 1960) meghatá-  
rozása szerint az egyes hangok azon tulajdonsága, mely alapján a hallgató képes két ugya-  
nolyan hangerejű és ugyanolyan alapfrekvenciájú hangot különbözőkként felismerni. Bár a  
hangszín elsődlegesen a hang spektrumának függvénye, függ a hang hullámformájától, a  
hangnyomástól, valamint a hang időszerkezetétől is (American Standards Association 1960). A  
hangszín tehát az, ami például elkülöníti egymástól a különböző hangszerek hangját – és mivel  
a hangszerek testének alakja vagy anyaga (szinte) sosem változik, ez a meghatározás a hang-  
szerek világában elegendő is. Más a helyzet azonban az emberi beszéddel – itt ugyanis a hang-

szín az egyes beszélők eltérő hangszínére és az egyes magánhangzók egymástól eltérő hangszínére is utalhat. Éppen ezért Gósy (1999) a jelenségek megnevezésének egyértelműsítése céljából megkülönbözteti a *hangszín* és a *hangszínezet* fogalmakat, előbbit a különböző magánhangzó-minőségekre, utóbbit pedig az egyéni beszédjellemzőkre értve. A jelen értekezésben ezt a terminológiai elválasztást követjük, így a magánhangzókat egymástól elkülönítő akusztikai-auditív sajátosságaira a *hangszín* terminussal, a beszélő hangjának vagy az éneklésnek a sajátos akusztikai „színezetére” pedig a *hangszínezet* terminussal utalunk.

A második, az értekezés kérdésfelvetései szempontjából alapvetően fontos fogalom a formáns –ez ugyanis bár egy nagyon gyakran használt, és alapvetőnek tekintett, mégsem egyértelmű terminusa a beszédtudománynak. A formáns első definícióját Gunnar Fant (1960) adta, eszerint a formánsok a hangspektrum csúcsai, vagyis a hangnyomás-idő függvény spektrumának energiamaximumai. Emellett azonban Fant (1960) megemlítette azt is, hogy a kimeneti energiamaximumok frekvenciahelyeihez közel eső, a toldalékcső átvitelére (*gain function* ’erősítés függvény’  $|T(f)|$ ) jellemző rezonanciafrekvenciákat a formánsoktól elkülönülten kell kezelni, annak ellenére, hogy ezek valóban sokszor egybesznek a kimenetben mérhető csúcsokkal (tehát a formáns kifejezés a rezonanciacsúcsokra is használható). Fant (1960) tehát konceptuálisan elkülönítette a toldalékcsőre jellemző sajátfrekvenciákat (ezeket rezonanciáknak nevezte), valamint a kimenetben mérhető energiacsúcsokat (melyekre a formáns elnevezést használta). A szakirodalom azonban nem egységes e hagyomány követésében, a formáns kifejezést ugyanis sok kutató minden esetben (tehát alaphangfrekvenciától függetlenül) az akusztikai kimenetben mérhető csúcsokra érti (lásd az Amerikai Akusztikai Társaság meghatározása [ANSI 1994], valamint pl. Benade 1976, Gósy 2004, Németh – Olaszky szerk.), míg mások a toldalékcső sajátrezonanciáinak megjelölésére használják (l. pl. Stevens–House 1961, Sundberg 1975). Az előbbi meghatározás értelmében definiált formánsok (azaz a spektrális energiamaximumok) alacsonyabb alaphangfrekvenciák esetében általában egybeesnek az utóbbi értelemben definiált formánsokkal (azaz az üregi sajátrezonanciákkal), de magasabb alaphangfrekvenciák esetében nem. Magasabb alaphangfrekvenciákon az előbbi definíció szerint meghatározott formáns elnevezés már nem értendő (és nem is érthető) a toldalékcsőre jellemző sajátrezonanciákra, hanem kizárólag az akusztikai kimenetben mérhető spektrális energiamaximumokra,

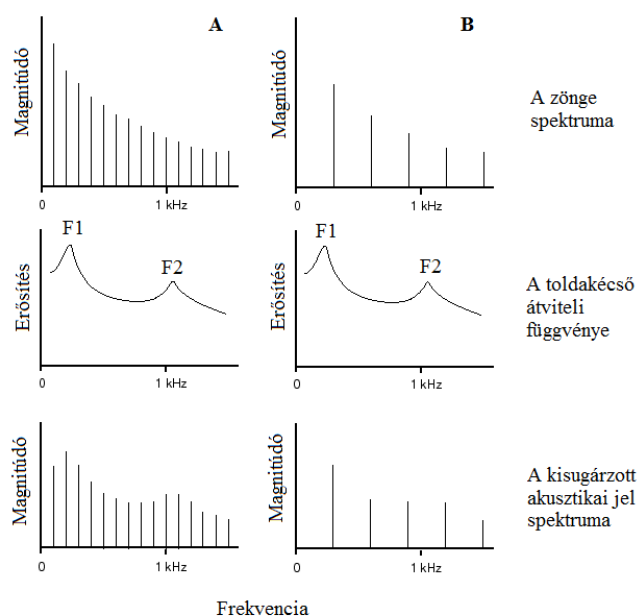


amivel egyidejűleg az artikulációval szoros összefüggést mutató rezonanciákra magas alapprofrekvencián már nem áll rendelkezésünkre elnevezés, azok az első meghatározás értelmében magas alapprofrekvencián gyakorlatilag értelmezhetetlenné válnak.

Mivel a formánst alapvetően az első definíció szerint meghatározó vizsgálatokban is gyakran látunk az artikulációs konfigurációra vonatkozó következtetéseket, valamint ezek a vizsgálatok is, úgy tűnik, alapvetően rezonanciaként utalnak a formánsokra, úgy látjuk helyesnek, ha a jelen értekezésben – melyben a magas alapprofrekvenciák kérdése igen fontos lesz –, a fent megfogalmazottak közül a második formánsdefiníciót használjuk. Eszerint tehát formánsnak elsősorban nem az akusztikai kimenetben mérhető spektrális energiamaximumokat, hanem az egyes hangzók ejtésekor az artikulációs konfigurációra jellemző átviteli függvény csúcsait, azaz a toldalékcso s ajátfrekvenciáit, rezonanciáit tekintjük. Ezek természetesen alacsonyabb alapprofrekvenciákon jobban megfeleltethetők egymásnak, magasabb alapprofrekvenciákon azonban nem, ezért a magas alapprofrekvenciákon tapasztalható artikulációs (és akusztikai) stratégiákra a spektrális maximumokból csak áttételesen következtethetünk (vagy fogalmazhatunk meg sejtéseket). A értekezés azon pontjain, ahol a formáns fogalmát ettől a meghatározástól eltérően (is) értelmezzük, és figyelmet szentelünk a percepció szempontjából fontos (akusztikailag is mérhető) spektrális maximumhelyeknek is, ezt megjegyzésként minden esetben fel-tüntetjük (vö. 2.1 fejezet).

Az alapprofrekvencia (és a felharmonikusok), valamint a formánsok között fennálló már korábban említett összefüggés miatt kellően mély alapprofrekvenciák (például az átlagos felnőtt beszélők beszédhangjának) esetében a zöngére jellemző nagy felhangsűrűség miatt úgy tekin-tethetjük, hogy a formánsok mint a toldalékcso rezonanciai helyei jó közelítéssel megállapíthatók a beszédhangok spektrumából, hiszen a mély alapprofrekvenciák esetében – amint láttuk – nagyobb számú felharmonikus esik egy-egy rezonancia hely sáv szélességébe (tehát több felharmonikus vesz részt a spektrális csúcsok kialakításában), így alacsonyabb alapprofrekvenciákon az átviteli függvény jobb frekvenciafelbontású lesz a beszédhangok akusztikai lenyomatában (vö. Wolfe et al. 2009). Az alapprofrekvencia emelésével azonban, ami a jelen értekezés kérdés-felvetéseinek egyik alapvető problémakörét képezi, a felharmonikusok egyre növekvő távol-sága miatt (vö. 6. és 7. ábra) a formánsok (mint rezonanciák) detektálhatósága, elérhetősége az

akusztikai kimenetben mind az akusztikai mérések, mind pedig az emberi beszédpercepció számára egyre csökken. A 7. ábra mély (A) és magas (B) alapfrekvenciájú hangok vonalas spektrumát ábrázolja. Jól látható, hogy míg alacsony alapfrekvencián az F1 és az F2 rezonanciák egyaránt jól kivehetők az akusztikai kimenetben, a magas alapfrekvencián sem az F1, sem pedig az F2 rezonancia frekvenciahelye nem detektálható nagy bizonyossággal (bár igaz, hogy az  $f_0$ , azaz a spektrum legalacsonyabb frekvenciaösszetevője nagy intenzitást mutat, tehát feltehetőleg erősített).



7. ábra: A hang keletkezése a forrás-szűrő modellben alacsony (A) és magas (B) alapfrekvenciák esetén: fent a gégénél létrejött a zöngespektruma, középen a szűrőként funkcionáló átviteli függvény, alul pedig a szűrt, a szájnál kisugárzott akusztikai jel (Epps et al. 1997: 1113 alapján)

A fentiek értelmében tehát azt mondhatjuk, hogy a toldalékcsőre jellemző sajátfrekvenciák vagy formánsok a magasabb alapfrekvenciák esetében már nem feltétlenül egyeznek meg az akusztikai kimenetben megjelenő energiamaximumokkal – a kimenet egyes akusztikai összetevőinek intenzitása inkább csak következtetni enged arra, hogy az egyes felharmonikusok be-

leesnek-e az egyes formánshelyek sávszélességébe. Így mind a formánsok mint rezonanciák frekvenciájára, mind pedig az azokat létrehozó artikulációs mozgásokra is egyre kisebb biztonsággal tudunk következtetni az alapfrekvencia emelésével.

A fent leírt jelenséget, mely tehát a toldalékcső átviteli függvényének az alapfrekvencia emelésével egyre alacsonyabb frekvencia-felbontását eredményezi, egyes kutatók a beszédhangok magas alapfrekvencián tapasztalható spektrális alul-mintavételezettségének nevezik (de Cheveigné–Kawahara 1999). Bár ez a terminus talán megtévesztő lehet, ha a mintavételezés fogalmát mereven a digitális jelfeldolgozásban használatos értelemben próbáljuk alkalmazni (vö. Szaszák 2010), mégis az analógia nagyon szemléletesen és egyszerűen ragadja meg a „mintavételezési frekvenciából” (azaz itt a felharmonikusok távolságából, illetve az alapfrekvenciából) fakadó felbontásbeli különbségeket az alapfrekvencia függvényében. Éppen ezért a magas alapfrekvenciájú hangok spektrális alul-mintavételezettségére a jelen értekezésben is hivatkozni fogunk, a formáns fogalmát pedig tehát a toldalékcső rezonanciáira alkalmazzuk.

#### 1.1.1.3. A magánhangzók artikulációs és akusztikai jellemzése

A magánhangzókat artikulációsan és akusztikailag is jellemezhetjük: az előbbi a magánhangzók létrehozását előidéző artikulációs mozgásoknak, az artikulátorok állásának, utóbbit pedig egyebek mellett – a fent említett csőmodell alapján – az artikulációval szorosan összefüggő formánsfrekvenciáknak a megadása jelenti. Az artikulációs és akusztikai meghatározás célja, hogy elkülönítsük és jellemezzük az egyes magánhangzó-minőségeket. A *magánhangzó-minőség* terminus technicus az ejtett magánhangzó auditív minőségét jelöli – az a címke, melyet a három alapvető, a magánhangzók ejtését jellemző paraméter határoz meg: az elől- vagy hátulképzettség, a nyitottság vagy függőleges nyelvhelyzet és az ajakműködés (Ashby 2011). Éppen ezért ezek azok a paraméterek, melyeket az artikulációs jellemzéskor megadunk. Az egyes paraméterek pedig szorosan összefüggnek a toldalékcsőre a magánhangzók képzésekor jellemző sajátfrekvenciákkal, formánsokkal, aminek eredményeképp a magánhangzók formánsai alkalmasak lesznek arra, hogy akusztikailag elkülönítsék és meghatározzák az egyes magánhangzó-minőségeket (vö. Fant 1960).

A magánhangzók képzésekor a hangszalagok (normál beszédmódban) mindig rezegnek<sup>2</sup>, az aktív és passzív artikulátorok, azaz a nyelv és a szájpad között létrejövő szűkület pedig megfelelően nagy ahhoz, hogy az artikulátorok ne érintkezzenek (tehát nem okoznak zárat, mint a zármássalhangzók pl. a /t/ esetében,), és hogy a szűkületen átáramló levegő ne okozzon turbulens zörejt (mint a résmássalhangzók, pl. az /s/ esetében). A zöngé jelenléte és az érintkezés vagy a zörej hiánya miatt a magánhangzókat szonoránsoknak (vö. pl. Gósy 2004, Ashby 2011) soroljuk be.

Bizonyos értelemben a magánhangzók csoportosításának legalapvetőbb dimenziója a nyíltság. E szerint megkülönböztetünk magas(abb), illetve alacsony(abb) nyelvháttal képzett beszédhangokat (vö. Stevens 1998). A nyíltságnak jellemzően négy fokozatát különítjük el, ezeket magyarul felső, középső, alsó és legalsó nyelvéllésnek nevezzük (Gósy 2004). Amint azt a terminológia is mutatja, a nyíltság dimenzióját sokan elsősorban a nyelvtest függőleges helyzetének feleltetik meg. Ugyanakkor a nyíltság kérdése az artikuláció, az akusztikum és a percepció figyelembevételével valójában ennél egy kicsit összetettebb, melyben nem csak a nyelvtest, de az állkapocsnyitás szöge is fontos szereppel bírhat – erről még a következő alfejezetben a magyar magánhangzók ismertetésekor valamivel bővebben is szólnunk.

Akusztikai tekintetben a magasabb nyelvéllés következményeként a létrejött hang F1 értéke alacsonyabb lesz, az alacsonyabb nyelvéllés következményeképp az F1 értéke magasabb. Ha a nyelvéllés foka magas, a garatüreg térfogata a neutrális nyelvhelyzethez képest nagyobb, mely esetben az F1 értékét elsősorban a létrejött szűkület mérete, hossza határozza meg. Az F1 értékének maximális lesüllyesztését az ajkak réses nyitása és az ajaknyílás keskenyítése, hosszítása révén érhetjük el, mely éppen az /i:/ hang artikulációjára jellemző – ez tehát az artikulációs és akusztikai magánhangzótér egyik sarokpontja (Stevens 1998).

Az állkapocs nyitása és a nyelvéllésfok csökkentése emeli az F1 értékét. A nyugalmi helyzethez képest az F1 növekedését okozza ugyanis, ha a nyelv lefelé és kicsit hátrafelé, a garat felé húzódik, melynek következtében a szűkület a szájüreg hátsó részében, a garatnál jön létre (de fontos, hogy ekkor a szűkület még nem elég kicsi ahhoz, hogy turbulens zörejt hozzon lét-

---

<sup>2</sup> Bár folyamatos beszédben a hangzók egymásra hatásának következtében előfordulhat a magánhangzók zöngétlenedése, a zöngétlen magánhangzókat a magyarban nem tekintjük önálló, fonémaszerű hangoknak.

re). Ezt az idézi elő, ha a nyelvtest elülső része lefelé mozdul, a hatást pedig erősíti, azaz az F1 értékét tovább növeli, ha megnövekszik az szájüreg elülső részének térfogata is, melyet az állkapocsnyitás szögének növelésével idézhetünk elő (Stevens 1998).

Az F1 értéke tehát összefoglalóan a nyelvtest függőleges helyzetével, valamint az állkapocs nyitásszögével függ össze, így a magánhangzók nyíltságát akusztikailag az F1, artikulációsan pedig a nyelvtest és/vagy az állkapocs nyitásszögével jellemezzük.

Az F2 értékére elsősorban a nyelv vízszintes helyzete hat, azaz az a körülmény, hogy a nyelvtest elől vagy hátul helyezkedik-e el a szájüregben. Általánosan a nyelvtest előretolulása az F2 értékének növekedését eredményezi, míg a nyelvtest hátrafelé pozicionálása az F2 értékének csökkenését okozza – igaz azonban az is, hogy az F2 maximuma a felső nyelvállásúak esetében magasabb lehet, mint az alsó nyelvállásúaknál (Stevens 1998). Stevens (1998) modelljének adatai alapján megállapítható, hogy a felső(bb) nyelvállású magánhangzók F2-értékének hatékony csökkentéséhez szükséges az ajakkerekítés is, így a legzártabb (legalacsonyabb F1-gyel képzett) és legalacsonyabb F2-vel képzett /u/ jelenti az artikulációs és akusztikai magánhangzótér egy másik sarokpontját (minimális F1- és F2-értékkel, legzártabb és egyúttal leghátrébb képzett hangként). Az ajakkerekítés hatása az F2 értékére azt a tendenciát is magyarázza, mely szerint „jelöltebbek” a hátul képzett réses hangzók, azaz gyakoribbak azok a nyelvek, melyekben csak kerekítéses hátul képzett magánhangzók fordulnak elő. A hátul képzett hangzók ajakkerekítéses ejtése ugyanis a fentiek alapján akusztikailag motiváltabbnak tűnik, hiszen ez teszi lehetővé a legnagyobb akusztikai kontrasztot az elől képzett hangokhoz képest.

Noha az alsó nyelvállásfokú hangok mozgástere az elől-hátulképzettség dimenziója mentén fizikailag kisebbnek tűnhet a felsőbb nyelvállású hangokénál, mégis lehetséges az elől- és hátulképzettség distinkciója, mégpedig úgy, hogy a létrejövő szűkület helye az F1 értékét csak minimálisan befolyásolja. Mégis, a legalsó nyelvállású elől képzett hangzók maximális F1- és F2- értékét az ajakréses ejtés tovább fokozhatja, ugyanis az ajakréses ejtés valamelyest emeli az F2 értékét (Stevens 1998).

A magánhangzók jellemzésének második artikulációs és akusztikai paramétere tehát a vízszintes nyelvhelyzet – ez alapvetően két vagy három fokozatú lehet, ti. az elől, a hátul vagy a

középen képzett hangok esetében –, valamint az F2, mely elsősorban az imént említett artikulációs dimenzióval mutat szoros összefüggést. Láttuk azonban azt is, hogy az F2 értékére valamelyest az ajkak réses vagy kerekítéses formálása is hatással van, hiszen a réses ejtés némileg emeli, az ajakkerekítéses ejtés pedig csökkenti az F2 értékét. Ennek értelmében az ajkak működése a harmadik, a magánhangzók jellemzésében általánosan megadott artikulációs paraméter, melyhez akusztikusan szintén az F2 értékét társíthatjuk.

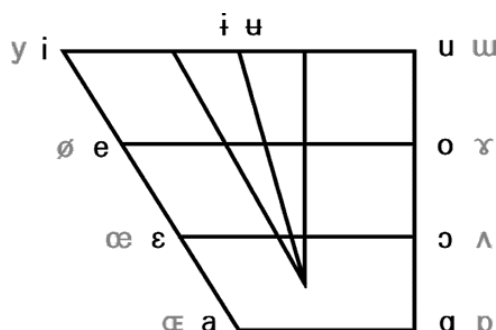
A fentiek rávilágítanak arra a korábban már említett tényre, hogy miért bír a toldalékcso első két sajátrezonanciája, azaz az első két formáns, az F1 és F2 kitüntetett jelentőséggel a beszédtudományban. Az F1 szoros összefüggése a magánhangzók nyíltságával, valamint az F2 szoros összefüggése a magánhangzók elől vagy hátul képzettségével ugyanis egyfelől könnyen (sőt sokszor félrevezetően könnyen) interpretálhatóvá teszi az első két formáns értékét, tehát a magánhangzók akusztikai jellemzőit artikulációs tekintetben, másfelől pedig ez az összefüggés a magánhangzó-minőségek kétféle (ha szigorúan nem is ekvivalens, de) szorosan összekapcsolható jellemzését teszi lehetővé az artikulációs és az akusztikai vetületben.

A teljesség kedvéért megemlítendő, hogy a fentiek mellett megkülönböztetünk még orális és nazoorális magánhangzókat (tehát a magánhangzók képzéséhez felvehetünk még egy artikulációs képzési jegyet), valamint monoftongus, diftongus és triftongus magánhangzókat is. Azonban ezek megkülönböztetése a magyar köznyelvi magánhangzókészlet szempontjából másodlagos, hiszen a készlet minden eleme orális és monoftongus beszédhang. A nazoorális magánhangzók ejtése során (az orálisokkal szemben) a lágy szájpad nyitva van (tehát nem zárja el a levegő útját az orrüreg felé), ezért az ejtés során a levegő nem csak a szájon, de az orron keresztül is távozik, ami sajátos hangszínt ad a beszédhangoknak. A diftongusok (és triftongusok) a monoftongusokkal szemben olyan magánhangzók, melyek képzési ideje alatt valamely, a magánhangzó minőségét meghatározó jegyben (pl. a nyelvállás fokában) egy (vagy a triftongusok esetében egymás után több) változás történik. A változás következtében pedig valójában kettős- (vagy hármas-) hangzót ejtünk, olyan módon, hogy a két (vagy három) kiejtett hang nem képezhet például önálló szótagot, így nem is számít önálló fonémaértékkel bíró hangnak sem – egyetlen fonémát deifinál (Gósy 2004).

Bár a fenti jellemzés következtében a magánhangzók rendszerezése ma már viszonylag egyszerűnek tűnhet, nem volt mindig ilyen kézenfekvő. Amiatt ugyanis, hogy a magánhangzók képzésekor az artikulátorokat csak közelítjük, de nem érintjük egymáshoz, illetve a közelítés hatására nem jön létre zörej, valójában a magánhangzók jellemzése a mássalhangzóknál jóval nehezebb kérdésnek bizonyult, két okból is. Egyfelől bár az előző alfejezetben említettük, hogy az egyes magánhangzók képzésekor a nyelv érintkezhet ugyan a nem mozgatható artikulátorokkal (pl. a nyelvperem a felső nyelvállású magánhangzók esetében hozzáér a felső fogsorhoz vö. Stevens 1998), mégis alapvetően beszéd közben nem kapunk jól megfigyelhető, pontos taktilis visszajelzéseket a magánhangzók létrehozását eredményező közelítés vagy szűkület létrejöttének helyéről, azaz a magánhangzó képzéshelyéről. Másfelől pedig a magánhangzók izolált ejtése statikus – szemben a mássalhangzókéval –, így kinesztetikus információk sem segítik a magánhangzók képzéshelyének azonosítását, miközben pedig tudjuk, a nyelv igen nagy varibilitással mozgatható és alakítható (Ashby 2011).

Daniel Jones (1922) a magánhangzók jellemzésének problémáját megoldandó megalkotta a kardinális magánhangzók rendszerét, azaz viszonyítási pontokat határozott meg a magánhangzók képzéséhez használt artikulációs és akusztikai térben. Jones (1922) pusztán kivételesen jó hallására támaszkodva képes volt gyakorlatilag a jelen fejezetben korábban megfogalmazott artikulációs, akusztikai és a következőkben érintendő percepciós sajátosságok megragadására, és bizonyos szélső értékeket képező hangzóminőségek lehorgonyzásával rendszerbe szedte az emberi beszédképzőszervekkel létrehozható magánhangzókat. A kardinális magánhangzók rendszere nem egy létező nyelv magánhangzóinak rendszere, hanem az artikulációs „munkaterület”, az artikulációs magánhangzótér felosztása egymástól – és ez egy igen fontos kitétel – *auditíve* egyenlő távolságokra lévő kategóriákra. Ilyen módon a kardinális magánhangzók inkább absztrakt rendszere jó (auditív) mérőeszköze lehet az egyes realiztikus magánhangzók-nak, tehát lehetővé teszi, hogy a beszédhangokat rendszerezzük, és a különböző nyelvek magánhangzóit egymással összehasonlítható módon jellemezzük. A rendszerben megadott magánhangzók rögzítettek, és mint ilyenek tehát maguk nem realiztikusak. Céljuk pusztán az, hogy sarokpontokat képezzenek, melyekhez képest az egyes nyelvek realiztikus beszédhangjai „lehorgonyozhatók” (vö. pl. Ashby 2011).

Az első rendszer nyolc magánhangzó-minőséget tartalmazott, melyet aztán Jones megduplázott az ajakműködés szerint – minden korábban kerekítéses hang mellé egy réses ejtésű párt rendelt és fordítva (l. 8. ábra). Érdekes módon a kardinális magánhangzók rendszerében a magánhangzók elkülönítése hallásalapú, mégis a magánhangzókat leíró terminológia alapvetően artikulációs tényeken és terminológián alapszik (a magánhangzókat tehát artikulációsan jellemezzük és különítjük el egymástól). A rendszert többek közt éppen emiatt számos kritika érte. Mivel azonban az artikuláció (nyelvhelyzet) és annak auditív eredménye, azaz a „magánhangzó-minőség” között erős korreláció tapasztalható, ezért ez a fajta megközelítés – normál beszédejtésben legalábbis – jó közelítéssel teszi lehetővé a magánhangzók jellemzését. Jól mutatja ezt az a tény, hogy a magánhangzókat máig e szerint a rendszer szerint adjuk meg. Az első bővítést aztán további is követte, így alakult ki a jelenleg is használatos szisztéma.



8. ábra: Az elsődleges (fekete) és másodlagos (szürke) kardinális magánhangzók rendszere (Forrás: <http://matters-phonetic.blogspot.hu/2011/02/romantic-cardinal-vowels.html>)

A kardinális magánhangzóknak a 8. ábrán bemutatott rendszere, azaz a magánhangzótrapéz a következőképpen olvasandó. A függőleges tengelyről a nyelvállás fokát vagy az állkapocsnyitás szögét olvashatjuk le: fent találhatók a felső nyelvállású, zárt magánhangzók, alul pedig az alsó nyelvállású, nyílt magánhangzók. A vízszintes tengely pedig a nyelv vízszintes pozícióját mutatja balra a palatális, azaz elől képzett, jobbra pedig a veláris vagy hátul képzett hangokkal. A vízszintes tengely mentén középtájon elkülönül a centrális vagy mediális kép-



zeshely is. A magánhangzópárok bal oldali tagja minden esetben ajakkerekítéssel képzett, labiális hang, míg a jobb oldali tag ajakréssel képzett, azaz illabiális.

Mivel a vízszintes és függőleges tengelyek az F1 és F2 értékével szoros összefüggésbe hozható artikulációs paramétereket ábrázolják, ezért a Jones (1922) rendszeréből kiindult, a nyelvhelyzetet is figyelembe vevő, és ilyen értelemben inkább az artikulációs magánhangzóteret leképező IPA szabvány szerinti ábrázolás (vö. IPA Handbook 1999) megközelítően jól reprodukálható az F1 és az F2 tengelyek mentén felvett akusztikai magánhangzótérrel. Bár ez utóbbi, tehát az akusztikai magánhangzótér nem feleltethető meg maradéktalanul az előbbinek, azaz az artikulációs magánhangzótérnek (például azért sem, mert az ajakkerekítés miatt alacsonyabb F2 az akusztikai magánhangzótérben így elkerülhetetlenül a hátulképzettség „számlájára íródna”), a kettő elkülönítésének figyelembevételével az akusztikai magánhangzótér ábrázolása és vizsgálata jó és gyakran használt eszköz a magánhangzók megvalósulásának elemzésében.

A fent bemutatottak alapján az egyes magánhangzó-minőségek tehát artikulációs és akusztikai tekintetben elkülönölni látszanak egymástól, hiszen a nyíltság és a vízszintes nyelvhelyzet más-más akusztikai szerkezetet, eltérő F1 és F2 értékeket eredményez. Ebből kiindulva hagyományosan szokás feltételezni, hogy az egyes magánhangzók felismerése is alapvetően ezeken a kitüntetett szerepű frekvenciaösszetevőkön múlik. A valóságban azonban az F1 és F2 értékek igen nagy variabilitást mutathatnak, és sokszor átfedésben is vannak az egyes magánhangzó-minőségek között (legalábbis a lineáris fizikai, azaz Hertz alapú ábrázolásban). A magánhangzók észlelése, azaz az egyes magánhangzó-minőségek azonosítása tehát korántsem triviális működések eredménye. Minderről a magyar magánhangzók a következő fejezetben bemutatott ismertetése után lesz szó.

#### **1.1.1.4. A magyar magánhangzók artikulációs és akusztikai jellemzése**

A magyarban a magánhangzók (artikulációs) jellemzéséhez hagyományosan (valamint az előző fejezetben bemutatottaknak is megfelelően) a nyelv nyugalmi pozíciójához viszonyított vízszintes és függőleges helyzetét adjuk meg, valamint az ajakaktivitást jellemezzük. Ezen

felül pedig, mivel a magyarban jelentésmegkülönböztető szerepe van, szokás megadni a magánhangzók nyelvi, azaz fonológiai időtartamát is (vö. Gósy 2004).

A nyelv függőleges helyzete szerint megkülönböztetünk felső, középső, alsó és legalsó nyelvvállásfokokat, a nyelv vízszintes helyzete szerint elöl és hátul képzett vagy másként palatális és veláris hangzókat, valamint középen képzett, azaz centrális vagy mediális hangzókat, az ajakműködés szerint elkülönítünk ajakréssel képzett réses vagy illabiális és ajakkerekítéssel képzett kerekített vagy labiális hangzókat, míg a nyelvi hosszúság tekintetében megkülönböztetjük a fonológiailag hosszú és rövid hangokat (l. Bolla 1995, Kassai, 1998, Gósy 2004, Mády 2008b).

A magyar magánhangzóknak a nyelvvállásfok szerinti táblázatos összefoglalását látjuk az 1. táblázatban. A táblázat csak a minden idézett leírásban egyaránt meglévő 14 magánhangzót tartalmazza (ezek a hangzók azok, melyek a leírások szerint nem kötődnek bizonyos beszédhelyzetekhez vagy nyelvváltozatokhoz), így nem szerepel benne a Bolla (1995) és Kassai (1998) által feltételezett zárt *ě*, azaz a rövid /e/, valamint a Gósy (2004) által feltételezett rövid /a/ és a svá (/ə/).

1. táblázat: A magyar (nyelvváltozatokhoz vagy beszéd módokhoz nem kötődő megjelenésű) magánhangzók nyelvvállásfok/nyíltság szerinti fonetikai\* besorolása a különböző szakirodalmi forrásokban (az első oszlopban a nyíltság mértékének az értekezésben is több helyütt hivatkozott számszerű kifejezését adtuk meg).

Nyíltsági fok	A nyelvvállás foka	Bolla (1995)	Kassai (1998)	Szende - IPA Handbook (1999)	Gósy (2004)	Mády (2008b)
1.	Felső	i i: yy: u u:	i i: y y: u u:	i i: y y: u u:	i i: y y: u u:	i i: y y: u u:
2.	Középső	e: ø ø: o o:	e: ø ø: o o:	e: ø ø: o o:	e: ø ø: o o:	e: ø ø: o o:
3.	Alsó	ɛ ɔ	ɛ ɔ	ɛ	ɛ ɔ	ɛ
4.	Legalsó	a:	a:	a: ɑ/ɒ	a:	a: ɒ

\* A fonológiai besorolás talán legnagyobb eltérése a fonetikai besoroláshoz képest a nyelvvállásfok tekintetében, hogy nem különböztet meg alsó és legalsó nyelvvállásfokokat, az /ɛ a: ɔ/ magánhangzókat egyszerűen alsónak ([–high][+low] jegyűnek) tekinti (vö. Siptár–Törkenczy 2000).

Noha az itt idézett szövegekben a magánhangzók nyíltsági korrelátumának alapvetően a nyelvállás fokát tekintik a szerzők (így ezt szerepeltettük a táblázat második oszlopában is), egyes leírások emellett (gyakorlatilag a nyelvállásfok szinonimájaként) megadják 1. a szájüreg nyíltsági fokát is az állkapocs nyitásszögében kifejezve, azaz (a nyelvállásfokok fenti felsorolásának megfelelően) rendre a zárt, félig zárt, nyílt és legnyíltabb állkapocsnyitást, valamint 2. az ajkak által képzett nyílás tágasságát is, mely értékei rendre: legszűkebb, szűk, tág, legtágabb (l. pl. Kassai 1998). A nemzetközi szakirodalomban ez a gyakorlat egyébként egyáltalán nem szokatlan, amint azt később látni fogjuk, Wood (1975) vagy akár Traunmüller (1981) is az állkapocsnyitás szögét jelöli meg, mint a magánhangzó nyíltságának elsődleges artikulációs korrelátuma (Wood 1975 szerint ehhez a nyelvállásfok mint a feszes-laza szembenállás artikulációs korrelátuma járul). Traunmüller (1981) ráadásul a magánhangzók percepciós nyíltsága kapcsán is a nyíltság, nyitásfok vagy nyíltsági fok kifejezéseknek fordítható *openness* és *degree of openness* terminusokat használja.

Traunmüller érvelése szerint a magánhangzók a nyíltság szerint észlelt distinktív jegyeinek (azaz a különböző nyitásfokoknak) a megragadása az egyszerűség és egyértelműség kedvéért szükségszerűen artikulációs terminusokkal (azaz a nyelvállás, az állkapocsnyitás vagy az ajaknyitás megadásával) kell, hogy éljen, azonban a legjobb fonológiai általánosíthatóságot az állkapocsnyitásszög megjelölése adja, hiszen a nyelvállás valójában a veláris és palatális, azaz az elől és hátul képzett hangok esetében némileg eltér, tehát variábilis (vö. Ladefoged 1967, idézi Traunmüller 1981, Gósy 2004). A jelen értekezésben, hogy elkerüljük az artikulációs, akusztikai vagy percepciós nyíltság megragadásának lehetséges dilemmáit, a magyar fonetikai hagyományban megszokott nyelvállásfok mellett alternatívaként hivatkozni fogunk a magánhangzók „nyíltságára”, valamint a nyíltság számszerűen kifejezett mértékére is (Traunmüller 1981 nyomán), mely nyíltsági fokokat az 1. táblázat első oszlopában tüntettük fel.

Az idézett források közti fő különbséget az *a* graféma által jelölt hang átírása és besorolása jelenti, azaz a szerzők nem egységesen feleltetik meg azt egy adott kardinális magánhangzónak, az *a*-t ugyanis az *á*-val (azaz /a:/-val) mint legelső nyelvállású hanggal szembeállítva több leírásban alsó nyelvállású hangként kategorizálják. A lehetséges megjelölések szerint tehát az *a* graféma által jelölt hang vagy ajakkerekítéssel képzett alsó nyelvállású /ɔ/, vagy ajakkerekí-

téssel képzett legalsó nyelvállású /ɒ/, vagy ajakréssel képzett legalsó nyelvállású /a/ magánhangzó. A hangzó vitatott besorolásával kapcsolatosan legújabban Mády (2008b) végzett artikulációs vizsgálatot elektromágneses artikulográf segítségével. Eredményei szerint bár az állkapocsnyitás kis mértékben valóban nagyobb volt az /a:/ esetében, mint a kérdéses hangzónál, a két magánhangzó esetében megfigyelt nyelvállásfok gyakorlatilag megegyezett, továbbá az *a* graféma jelölte magánhangzó esetében intenzív ajakkerekítés volt megfigyelhető. Mády következtetése szerint ezek az eredmények a többi magánhangzó ejtési mintázataihoz is viszonyítva, valamint korábbi vizsgálatok tanulságai alapján az *a* hang legalsó, az /a:/-val egyező nyelvállásfokú képzésére utalnak. A szerző érvelésében Wood (1975) munkáját idézi, akinek megállapítása szerint az ajakkerekítéses hangzók a szűkebb ajaknyílás következtében természetesen kisebb állkapocsnyítással is képzettek az azonos nyelvállásfokú párjaiknál, ennek megfelelően pusztán az állkapocsnyításban tapasztalt különbségek nem jelentik, hogy az adott magánhangzókat nyíltság vagy nyelvállásfok tekintetében különbözőnek kellene tekinteni. A jelen értekezésben ezen eredmények értelmében az *a* grafémával jelölt beszédhangot /ɒ/-ként jelöljük meg (tehát a Szende 1999 és Mády 2008b által javasolt besorolást alkalmazzuk).

A magyar magánhangzók leírása a nyelv vízszintes helyzete szempontjából nagyjából egységesnek mondható a szakirodalomban. A fent idézett források a veláris hangzók közé sorolják az /u u: o o: ɒ a:/ magánhangzókat, míg a palatálisok közé az /i i: y y: e: ε ø ø:/ magánhangzókat. A rendszer egyetlen valamelyest bizonytalannak tekinthető pontját az /a:/ és annak státusza képezi, mert bár az /a:/ akusztikailag centrális vagy mediális megvalósulást mutat – erre reflektál például Bolla (1995), Szende (1999) vagy Gósy (2012) is – hagyományosan egyértelműen a velárisokhoz sorolják (mely megfontolás legfőbb oka feltehetően a fonológiában keresendő, az /a:/ ugyanis például a magánhangzó-harmóniában veláris hangként viselkedik, valamint a veláris /ɒ/-val váltakozva vesz részt a fonológiai alternációkban is vö. pl. Siptár–Törkenczy 2000, Siptár 2014).

Ajakműködés tekintetében viszonylag nagy az egyezés Bolla (1995), Kassai (1998), Szende (1999) és Gósy (2004) között. Bár a szerzők mindegyike reflektál rá, hogy az alaptípusoknak számító ajakréses és ajakkerekítéses ejtés nyelvállásfokokként az állkapocsnyítás mértékétől függően változó megjelenést, változó ajakformát, elérő mértékű ajaknyitást mutat, a magán-

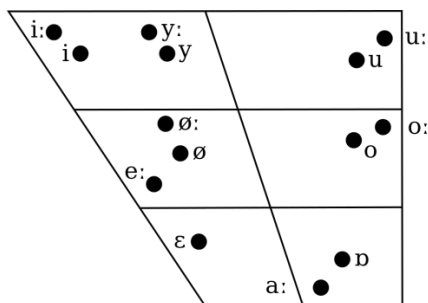
hangzókat egyértelműen besorolják az ajakkerekítéses és ajakréses csoportokba. A felosztás a következőképpen alakul. Ajakkerekítéses vagy labiális hangzók: /u u: o o: ɒ y y: ø ø:/, valamint ajakréses vagy illabiális hangzók: /i i: e: ε a:/. Ahogyan arra Bolla (1995) is utal, az alsó vagy legalsó nyelvéllésfokon ejtett labiális és illabiális magánhangzók közti különbség kisebbnek vagy legalábbis erősen más jellegűnek mutatkozhat, mint a felsőbb nyelvéllések esetében. Valószínűleg ez eredményezi azt is, hogy például Szende (1999) a vitatott minőségű *a* graféma által jelölt hangot elsődlegesen nem ajakkerekítéses /ɒ/, hanem ajakréses /a/ hangként írja át, majd megjegyzi, hogy mivel a hangzó valójában mutat némi kerekítést, így akár az /ɒ/-ként való átírása is lehetséges lenne. A jelen értekezésben, amint azt már fentebb is említettük, az *a* graféma által jelölt hang megjelölésére az ajakkerekítéses /ɒ/-t használjuk.

A magánhangzó-minőséget alapvetően meghatározó hangszínen (azaz az eddig taglalt három paraméteren) felül a magánhangzók jellemezhetők a nyelvi, azaz fonológiai hosszúságukkal is – különösképp fontos ez a magyarban, ahol a fonológiai hosszúság jelentésmegkülönböztető szerepű (vö. pl. Siptár–Törkenczy 2000). Ennek megfelelően a magánhangzók (az 1. táblázatban is szemléltetett módon) a fonológiai hosszúság szerint párokba rendezhetők. Amennyiben csak a fent idézett leírásokban egyaránt megjelölt (azaz a táblázatban feltüntetett) magánhangzókat tekintjük, a 14 magánhangzó 7 fonológiai hosszúság tekintetében opponáló pártba állítható. Bár az /e:/ és /ε/, valamint az /ɒ/ és /a:/ a hangszínük tekintetében jelentősen eltér, mégis fonológiai tekintetben pároknak tekinthetjük őket, elsősorban azért, mert nyelvi viselkedésüket tekintve a párok tagjai a csak a nyelvi időtartamukban eltérő párokhoz (pl. az /i i:/-hez) hasonlóan szisztematikusan (regulárisan) váltakoznak, azaz alternálnak (pl. a *víz* – *vizek*, *kéz* – *kezek*, *nyár* – *nyarak* alakokban vö. Siptár–Törkenczy 2000). Ugyanakkor éppen a hangszínbeli különbséget szem előtt tartva az utóbb említett beszédhangokat rövid vagy hosszú pár nélküli hangzóknak tekinti például Gósy (2004), az ő fonetikai szempontú rendszerezésében ugyanis csak a hangszínbeli különbséget nem mutató beszédhangok rendezhetők egymással hosszú-rövid párokba.

A jelen értekezés szempontjából kiemelendő, hogy a hangszínbeli és időtartambeli különbségeket is figyelembe véve úgy tekinthetjük, hogy a magyarban mindösszesen 9-féle magánhangzó-minőség különül el, melyek (a fonológiai párok hosszú tagjával szemléltetve) a követ-

kezők: /i: y: u: e: ø: o: ε a: ɒ/. Az ebből a felsorolásból kimaradt felső és középső nyelvvállású rövid beszédhangok minősége ugyanis (elméletileg) megegyezik a megfelelő hosszú pár minőségével, és a rövid-hosszú párok közti különbséget – ismét csak elsősorban elméletileg – csak a nyelvi időtartam adja. Valójában természetesen a helyzet nem ennyire egyszerű, ugyanis a hosszú magánhangzók hangszínükben (a nyelvvállás fokától is erősen függő módon) többé vagy kevésbé minden esetben eltérnek rövid párjaiktól (vö. Kassai 1998, Gósy 2004). Ez a jelenség (és az abban rejlő fonológiai is releváns probléma) jól megragadható a H & H elmélet keretei között (az elmélettel kapcsolatosan l. a jelen értekezés *1.1.2.1 alfejezetét*). A H & H elmélet értelmében a rövid-hosszú párok esetében tapasztalt hangszínbeli különbségek elvben a következő két ok valamelyikére vezethetők vissza (vö. Mády 2008b). 1. a magánhangzópárok célkonfigurációja megegyezik, a hangzók közti hangszínbeli különbség pusztán az alulkonfiguráltság vagy más szóval a célalulmúlás eredménye, hiszen a rövid magánhangzók képzésére fordított rövidebb idő miatt a beszélő nem képes elérni a hosszú magánhangzókra jellemző konfigurációt. Ez a feltételezés akkor valószínűsíthető, ha a rövid hangzók centralizáltabbak hosszú párjuknál, azaz a hosszúaknál valamivel jobban a magánhangzótér közepe felé tolódnak. Ez a középső és felső nyelvvállású rövid magánhangzók akusztikai adatai alapján általában sejthető tendencia (vö 9. ábra). 2. a rövid-hosszú magánhangzópárok között tapasztalt hangszínbeli különbség nem célalulmúlás eredménye, hanem annak a lenyomata, hogy a pár tagjainak célkonfigurációja eredendően eltér. Bár ezt eddig artikulációs vizsgálatokkal még nem igazolták, a magyar esetében feltehető, hogy a felső és középső nyelvvállású rövid magánhangzók artikulációs célkonfigurációja megegyezik a hosszú párjukéval (azaz a középső és felső nyelvvállású rövid és hosszú magánhangzók közti hangszínbeli különbség mindösszesen az artikulációra fordított idő következménye, vö. Mády 2008b). Éppen ezért a jelen értekezés erejéig magunk is elfogadottnak tekintjük azt a feltételezést, hogy a középső és felső nyelvvállású magánhangzók rövid párjai csak az alulkonfiguráltság miatt különböznek hangszínükben a megfelelő hosszú pártól, ezáltal azt is feltételezzük, hogy a magyarban 9-féle magánhangzó-minőséget különböztethetünk meg.

Az értekezésben feltételezett magánhangzókat a 9. ábra összesíti az IPA nemzetközi szabványának (IPA Handbook 1999) megfelelő, trapézos elrendezésben, mely az absztraktabb jellemzőket, valamint az artikulációs és az akusztikai sajátságokat együttesen ábrázolja.



9. ábra: A magyar magánhangzók az értekezésben feltételezett rendszere a hagyományos, trapéz formájú elrendezésben (Forrás: Wikipedia Commons, vö. Szende 1994: 92)

A magánhangzó-trapéz horizontális dimenziója az előző fejezetben már bemutatott módon a nyelv vízszintes irányú helyzetével függ össze: bal oldalon az elől képzett, míg jobb oldalon a hátul képzett hangokat találjuk (az oldalakat a centrális képzéshelyet jelölő függőleges vonal osztja ketté). A tér függőleges irányú dimenziója pedig, ismét csak a Jones (1922) rendszere kapcsán már taglaltaknak megfelelően a nyelv függőleges helyzetére utal: legfelül a felső nyelvállású/zárt/első nyíltsági fokú magánhangzókat, legalul pedig a legalsó nyelvállású/legnyíltabb/negyedik nyíltsági fokú magánhangzókat találjuk. Egy adott oldalon (tehát a palatális vagy veláris csoportokon belül) egy adott nyelvállásfokon bal kéz felé találjuk az ajakréses, míg jobb kéz felé az ajakkerekítéses hangzókat.

### 1.1.2. A magánhangzók percepciója

A pszicholingvisztikai modellekben a nyelvi feldolgozás legkisebb egységeként általában a hangszintű, más szóval szegmentális szintet említik. A legtöbb modellben ehhez a szinthez tartozik egy absztraktabb (fonológiai) és egy, a fizikai valósághoz közelebb álló (fonetikai) szint. Az ezekhez a szintekhez tartozó elemzési folyamatokat nevezzük átfogóan beszédészlelési

vagy beszédpercepciók folyamatoknak – míg a beszéd magasabb szintű egységeinek, a szavak, mondatok, megnyilatkozások feldolgozását is felölelő folyamatokat átfogóan a beszédmegértés folyamatának nevezzük (vö. Gósy 2004).

Bár a hangok absztrakt formája igen korlátozott számú, tehát az egyes nyelvek csak meghatározott számú fonémával gazdálkodnak, a hangok a fizikai valóság, azaz a fonetika szintjén igen változatos megjelenésűek, azaz variábilisak lehetnek – például attól függően, hogy milyen életkorú (azaz milyen toldalékcsoportokkal rendelkező) beszélő beszédében, milyen hangkörnyezetben jelennek meg, illetve hogy gyors vagy lassú beszédben, hangsúlyos vagy hangsúlytalan szótagban találhatók-e.

A magánhangzók esetében jelenleg a nemzetközi fonetikai ábécé, az IPA, huszonnyolc fonémát különböztet meg (l. IPA 2005). A beszédhangok azonosítása tehát azt jelenti, hogy az emberi beszédfeldolgozás valamilyen módon a számtalan lehetséges hangzómegvalósulás egy-egy példányát a kisszámú absztrakt és invariáns egységekhez, a fonémákhoz társítja. A feltételezések szerint ez a mechanizmus az invariáns absztrakciók társításához kiszűri a változatos megvalósulások szisztematikus, de fonetikailag-fonológiai nem distinktív jegyeit (amelyek felismerése, úgymond, nem segítené hozzá a hallgatót a beszédhang beazonosításához), és kiemeli, felismeri az invariáns jegyeket, melyek a fonetikai-fonológiai kategóriákat meghatározzák. Ezt a folyamatot, az itt megfogalmazott meghatározásból következő módon a szakirodalom gyakran normalizálásnak is nevezi (Neary 1989). Neary (1989) „normalizálendő” hatásokként elsősorban a beszélő- valamint a kontextusfüggő hatásokat említi, tehát megkülönbözteti a beszédhangoknak a beszélő egyéni ejtési és fiziológiai adottságaiból adódó, valamint a beszédhangnak a hangkörnyezetből fakadó (például a szomszédos mássalhangzó képzéshelye miatt jellemző) variabilitását. A beszédpercepciók elméletek (és vizsgálatok) fókuszában az itt említett normalizációs folyamat áll, valamint az a kérdés, hogy melyek azok az invariáns jegyek, amelyek felismerési kulcsokként szolgálnak egy-egy beszédhang kategóriába sorolásához.

A beszéd feldolgozása, ahogyan azt már említettük, nem áll meg a hangszintű elemzésnél – a beszédmegértés teljes folyamatának, és az ahhoz kapcsolódó elméleteknek az ismertetésére azonban a jelen munka keretei között nincs módunk. Mivel a jelen vizsgálatokban elsősorban



csak a szegmentális szintű beszédhang-azonosítással foglalkozunk, a jelen fejezet elméleti bevezetője is csak ennek kérdéseire fókuszál.

Mielőtt ismertetnénk a magánhangzók percepciójával kapcsolatos néhány fontosabb kutatási eredményt, röviden bemutatunk néhány, a beszédhangok percepciójával kapcsolatos alapvető elméletet – nemcsak azért, hogy megteremtsük a magánhangzókra vonatkozó eredmények értelmezéséhez szükséges keretet, hanem azért is, hogy szemléltessük a beszédhang-azonosítással kapcsolatos eddigi ismereteink sokféleségét – és bizonytalanságait is.

#### 1.1.2.1. Beszédpercepció elméletek

A beszédpercepció modelleket első közelítésben három nagy csoportba oszthatjuk – a három csoport a szerint oszlik meg, hogy az egyes percepció elméletek miben vélik biztosított-nak a beszédészleléshez szükségesnek tartott invarianciát. Az első a) csoportba tartozó modellek az invarianciát a magánhangzók artikulációjában, a második b) csoport elméletei a magánhangzók akusztikumában, míg a harmadik c) csoportba tartozó elméletek az észlelőben, az észlelő percepció mechanizmusaiban rejlőnek feltételezik<sup>3</sup> (vö. Lindblom et al. 1991).

a) Az első csoportba tartozó egyik legismertebb elmélet a beszédészlelés motoros elmélete (motor theory of speech perception). A modell elsősorban Liberman nevéhez és munkásságához (valamint a Haskins Laboratories intézményhez) kötődik, az elmélet kifejtését (vagyis annak javított változatát) Liberman és Mattingly publikálta 1985-ben. A motoros elmélet szerint a beszéd észleléséhez a kogníció egy szerkezetileg különálló egységet, modult tart fenn, melynek egyik feladata, hogy a beszélők által ejtett variábilis magánhangzó-megvalósulásokból dekódolja a hallgató számára a beszédhang létrehozásakor a beszélő toldalékcsovéra jellemző artikulációs konfigurációt, illetve artikulációs gesztusokat. A motoros elmélet azt feltételezi, hogy a beszéd során létrejövő akusztikai jel transzformációján keresztül a hallgató valójában az artikulációs információt detektálja, és ennek megfelelően az agy sem a hangok akusztikai lenyomatát, hanem a beszédhangok ejtéséhez szükséges motoros parancsokat (artikulációs

---

<sup>3</sup> Lindblom (1991) egyébként az első két elméletet egy nagyobb csoportba, a jelalapú elméletek csoportbába sorolja, míg az észlelő percepció mechanizmusaiban rejlő invarianciát feltételező, azaz saját H&H elméletét nem jelalapúnak nevezi.

gesztusokat) reprezentálja (mégpedig invariáns módon). A tüörkneuronok felfedezésével a kilencvenes években ez az elmélet új erőre kapott. A tükörneuronok olyan agyi idegsejtek, melyek mind egy adott cselekvés végrehajtása (pl. egy tárgy megragadása), mind pedig a cselekvés megfigyelése (a tárgy megragadásának látványa) közben „tüzelnek”<sup>4</sup>. Noha ez a megfigyelés önmagában csak annyit jelent, hogy motoros területek aktiválódnak a percepció folyamat közben (tehát a szenzoros és motoros működéseket egyfajta korreláció jellemzi), mégis gyakorta kauzalitásként értelmezi a szakirodalom, és azt állítja, hogy a motoros aktiváció célja az észlelési mechanizmus segítése. Bár a felfedezés erős empirikus bizonyítéknak tűnik a motoros teória helytállóságára, a szenzoros, azaz percepció működések közbeni motoros aktivációt, azaz a tükörneuronok működését egyes vizsgálatok a fentiekől eltérően magyarázzák. Hickok (2010) szerint a tükörneuronok működése elsősorban csak a szenzoros és motoros területek asszociációját mutatja, nem pedig azt, hogy a motoros aktivizáció a percepció szükséges része lenne (hiszen a korreláció nem értelmezhető törvényszerűen kauzalitási, oksági viszonyként). A Hickok által javasolt értelmezés szerint a percepció közbeni motoros aktiváció valójában az észlelés közben meginduló preparáció korrelátuma, azaz annak a lenyomata, hogy a megfigyelő a megfigyelt cselekvésre válaszüreakciót (például a beszéd esetében válaszümegszólalást) tervez. Bár a motoros teória állításai sok szempontból vitathatók, és az elméletet sok kritika éri, mégis számos jelenség (pl. a koartikuláció) megragadására sok szempontból alkalmasnak bizonyul, így ez a mai napig is használt beszédpercepció modellek egyike.

b) Az akusztikai (és bizonyos értelemben a percepció) invarianciára alapozó elméletek legismertebbje az elsősorban Stevens nevéhez (és a Massachusetts Institute of Technology, MIT intézményhez) kötődő kvantális elmélet. Stevens (1998, 1999) szerint az artikuláció és az akusztikum közti összefüggés nem lineáris, hanem ugrásszerű, azaz kvantális, tehát az akusztikai jel bizonyos mértékű és jellegű artikulációs eltérésekre érzéketlenebb, míg mások nagyobb változásokat okoznak benne. A kvantális elmélet tehát kategóriákat, valamint kategóriális észlelést feltételez, ami – ilyen módon – megkönnyíti az észlelés folyamatát. Az elmélet

---

<sup>4</sup> Azaz aktivációt mutatnak, reagálnak.

számos további beszédpercepció elképzeléssel, kísérleti eredménnyel kapcsolatba hozható (l. pl. Chistovich és Lublinskaja 1979 eredményei, melyről részletesebben a következő alfejezetben lesz szó), valamint abból számos további percepció teória is kinőtte magát. Egy ilyen elmélet például a LAFF (Lexical Access from Features, 'lexikai hozzáférés jegyek alapján'), mely szerint a hallgató az észlelés során nem fonémákat azonosít, hanem (bináris) distinktív (viszonylag egyszerű akusztikai korrelátumokkal rendelkező) jegyeket különböztet meg, a beszédjelet, a szavakat tehát e jegyek sorozataként észleli. Jegyen az elmélet képviselői nem a beszédjel hullámformájában egyértelműen elhatárolható szakaszokat vagy paramétereket értenek, hanem absztrakt egységeket: egy példa a [+high] azaz [+felső] jegy, ami a függőleges nyelvhelyzetre utal. Az egyes szavak felismerését, azaz aktivációjukat a LAFF elmélet szerint a mentális lexikonban az facilitálja, hogy a detektált jegysorozat az észlelő folyamatosan összeveti a mentális lexikonban tárolt mintázatokkal, majd a legkisebb különbség elvén „dönt”, és a szó aktiválódik (Stevens 1998, Stevens 2002). A LAFF előnye, hogy mivel az észlelést a fonémák azonosítása helyett a jegysorozatok azonosításában fogalmazza meg, tudja kezelni azt is, hogy az egyes jegyek akár időbeni átfedéssel is megjelenhetnek, tehát az elmélet alkalmas a koartikulációs folyamatok megragadására is. Ráadásul a modell kísérletes úton jól tesztelhető – bár az is igaz, hogy a modell állításaival kapcsolatos empirikus vizsgálatok szinte kizárólag csak mesterségesen (azaz laborban előhívott) beszédre alapulnak, így kérdéses, hogy vajon a spontán beszédben is érvényesek-e (Mády 2008a).

c) A harmadik, az invarianciát a hallgatóban feltételező irányzat híres képviselője Björn Lindblom, aki kutatási eredményeire támaszkodva megfogalmazta a H&H elméletet. Lindblom (1990) úgy véli, az állandóság nem az akusztikai vagy artikulációs jelben, hanem a hallgatóban keresendő, ugyanis az állandóság nem akusztikai vagy artikulációs jellemző, hanem a beszéd reprezentációjának tulajdonsága, a hallgató pedig nem „üres doboz”, ugyanis számos, az aktuális beszédjelen kívüli információ is a rendelkezésére áll. A H&H név a *hyper-* és *hypospeech* kifejezések rövidítéséből áll elő, melyek a fonetikai információban gazdag és az információhiányos jel koncepcióit fedik (egy kontinuum két végpontjaként). Lindblom elmélete szerint a beszélő a beszéd interakciós helyzetében az információgazdagság és információhiány skálája mentén rendelkezésére álló lehetséges megvalósítási formák közül választ

annak fényében, hogy mit feltételez a hallgató számára (a jeltől függetlenül) elérhető információról. Lindblom szerint a beszéd leképezésére szolgáló komplex reprezentációinkat azon jelekből konstruáljuk a nyelvelsajátítás során, melyeket a beszéd fonetikai és nem fonetikai jegyeiből ismerjük fel annak rendszerszerű változatossága révén. Ebből pedig az következik, hogy az állandóság valójában a reprezentáció tulajdonsága, és (paradoxnak tűnő módon) éppen a variabilitásban rejlik. Az elmélet fontos következtetése, hogy a variabilitás tehát nem a beszédfolyamat másodlagos vagy mellékes, hanem lényegi, esszenciális jellemzője. Az elmélet fontos alapfogalma a célkonfiguráció, melyen az egyes beszédhangok ideálizált artikulációs és akusztikai mintáját értjük, valamint a célalulmúlás, mely fogalom pedig a célkonfiguráció el nem érésének esetét jelöli (ezt látjuk például a magyar felső és középső nyelvválású rövid magánhangzók esetében, vö. Mády 2008b). Végezetül pedig kiemelt szerepet kap a modellben a kontraszt, ugyanis Lindblom javaslata szerint a kontraszt és a kontraszt erősítése szolgál a célalulmúlás révén „alulspecifikáltabb” beszédhangok közti különbségek megtartásának eszközéül (Lindblom 1990, 1992). A magánhangzók esetében a kontraszt erősítésének két prototipikus példája a túlartikulált beszédben nagyobb magánhangzótér (vö. pl. Wassink et al. 2007), vagy a fonológiai hosszúsági oppozíció erősítése szélsőségesebb fizikai időtartamok (általában nagyobb időtartamú hosszú magánhangzók) segítségével (vö. pl. Tajima et al. 2013).

#### **1.1.2.2. A magánhangzó-azonosítás kísérletes eredményei és a magánhangzószerű hangok feldolgozása a percepció rendszerben**

A magánhangzók (és magánhangzószerű hangok) azonosításáról máig viszonylag keveset tudunk, azonosításukról, az ahhoz szükséges (akusztikai) kulcsokról számos elmélet és kísérleti eredmény létezik. Mivel a magánhangzók legkarakterisztikusabb jegyei a hangzóspektrum prominenciái, azaz formánsai, így közvetve vagy közvetlenül a legtöbb elmélet ezekre épít. Már az első, a magánhangzók akusztikai szerkezetét szisztematikusan vizsgáló kísérlet is összefüggéseket talált a magánhangzó-spektrum prominenciái és a magánhangzók észlelése között (Peterson–Barney 1952). Ezen is alapulva általánosan véve úgy véljük (és a legtöbb tankönyvben, kézikönyvben is azt találjuk), hogy a magánhangzók azonosítása szempontjából

legfontosabb kulcs a magánhangzó első két formánsa (l. pl. Gósy 2004). Azt azonban, hogy az első két formánssal kapcsolatosan pontosan milyen információt hasznosít a beszédpercepció, már nem fogalmazhatjuk meg ilyen egyszerűen.

A jelen alfejezetben két kísérletileg alátámasztott elméletet ismertetünk, valamint egy további, az ismertett elméletekhez szorosan nem tartozó kutatási irányt, melynek az eredményeit a jelen értekezésben is hasznosítjuk. Az egyik elmélet a környéki hallórendszer működésével, azon belül is a hallóidegekkel, és azok tüzelési mintázataival hozza összefüggésbe a magánhangzók azonosítását. A másik elmélet a spektrális prominenciák egymáshoz viszonyított, érzeti összefüggéseire alapoz. Az ezek után bemutatott eredmények arra világítanak rá, hogy a magánhangzók felismerésében nem csak a magánhangzó középső, viszonylag állandó (stacioner) szakasza játszik (vagy játszhat) szerepet, de a magánhangzók koartikulációjából származó dinamikus akusztikai információk is, melyek az adott magánhangzó és a szomszédos hangok egymás utáni ejtéséből származó, a fonémák határian átívelő, egymással átfedésben lévő artikulációs gesztusok lenyomatai.

Békésy (1960) kutatásai óta tudható, hogy a kezdeti akusztikai elemzés már a fülben, az abban található csigában lezajlik, mégpedig az alaphártya működése révén. Az alaphártya ugyanis frekvenciaérzékeny, az alacsonyabb és magasabb frekvenciákra más-más részei reagálnak, mégpedig fizikai kimozdulással. Ezt a kimozdulást az alaphártyához kapcsolódó szőrsejtek felfogják, és idegi impulzussá alakítják, melyet a hallóidegek továbbítanak a központi idegrendszerbe, azaz az agyba. A hallóidegek azonban szintén frekvenciaérzékenyek, és saját karakterisztikus frekvenciával rendelkeznek. Ez azt jelenti, hogy minden idegrost egy bizonyos hangnyomás- és frekvenciatartományon tüzel, és van egy tartomány, ahol a legérzékenyebb (ahol a válaszhoz szükséges küszöbérték a legalacsonyabb) – ez adja az idegrost hangolási görbáját, valamint karakterisztikus frekvenciáját (Stevens 1998, Vicsi 2010a).

Az idegrostok tüzelési sajátosságain alapulva Sruлович és Goldstein (1983) egy olyan elméletet dolgozott ki, mely szerint a komplex hangok, például a magánhangzók, feldolgozásakor az idegek tüzelési mintázata az idegsejtek karakterisztikus frekvenciáinak függvényében szinkronizálódik a hangok spektrumának maximumaival (a magánhangzók esetében a formánssokkal), a komplex hangok feldolgozása tehát a spektrális maximumok és az idegsejtek saját-

frekvenciáinak összefüggéseitől függ. Későbbi pszichofizikai kísérletekben (l. pl. Delgutte–Kiang 1984) ezt az elméletet a magánhangzók azonosításával és kategorizálásával kapcsolatosan is vizsgálták. A vizsgálatok tanúsága szerint az idegi kisülések mintázatában látható, hogy azok szinkronizálódnak a magánhangzóspektrumok F1- és F2-prominenciáival – míg más, nem beszéd típusú jel (például valamilyen szélessávú zaj esetén) az idegrostok a saját karakterisztikus frekvenciájukon tüzelnek. Delgutte és Kiang (1984) tehát egyérszt kimutatta, hogy a környéki idegrendszer azonosítja a magánhangzóspektrum prominenciáit, azaz a formánsokat, másrészt pedig kimutatta azt is, hogy (az előbbieknél is megfelelően) az idegrostok tüzelési mintázatai eltérőek azokra a magánhangzókra, amelyek a zárt-nyitott, illetve az elől-hátulképzettség dimenziók más-más pontján helyezkednek el. Ez azért lehetséges, mert – mint tudjuk – az elől képzett hangzók F2-értékei magasabbak a hátulképzettekénél, valamint a felső nyelvállású, zárt ejtésű hangok F1-értékei alacsonyabbak az alsó nyelvállású, nyílt hangzókénál, tehát az  $f_0$ , az F1 és az F2 viszonya hangzóspezifikusan változik. A szerzők a vizsgálatukban azt találták, hogy ha a magánhangzóspektrum prominenciái megfelelően közel vannak egymáshoz (az  $f_0$ -F1 esetében ez a zárt képzésű magánhangzókra, míg az F1-F2 esetében ez a hátul képzett magánhangzókra igaz), a prominenciák frekvenciái közötti szörsejtek nem a saját karakterisztikus frekvenciájukon, hanem valamely hozzájuk frekvenciában közel eső prominencia frekvenciáján tüzelnek. Akkor viszont, ha a spektrum prominenciái a kritikus távolságnál nagyobb távolságra vannak egymástól, a két prominencia közti, köztes frekvenciákra érzékeny szörsejtek a saját frekvenciájukon tüzelnek, tehát az adott prominenciákat (az  $f_0$ -t az F1-től, illetve az F1-t az F2-től) elkülönítve érzékeljük. Mindez tehát azt jelenti, hogy bizonyos prominenciák (vagy prominenciapárok) közti viszonyok elkülönítik a zárt, felső nyelvállású hangzókat a nyílt, alsó nyelvállású hangzóktól, valamint a veláris hangzókat a palatálisoktól. Delgutte és Kiang (1984) magyarázata szerint az előbbi, azaz a zártság érzetét az  $f_0$  és az F1 távolsága, míg az utóbbit, azaz az elől-hátulképzettség érzetét az F1 és F2 távolsága befolyásolja.

Összefoglalóan tehát elmondható, hogy a magánhangzóspektrum és az idegrostok tüzelési sajátosságainak összefüggéseire alapuló vizsgálatok tanúsága szerint az elől és hátul képzett, valamint a zárt, felső nyelvállású és a nyitott, alsó nyelvállású magánhangzók már a környéki

feldolgozási folyamatokban természetes kategóriákként különülnek el egymástól. Mindez azonban ilyen tisztán csak a legalsó és legfelső nyelvállású hangzókra igaz, a köztes hangzók esetében ugyanis az idegek az alsó régiókban vagy az  $f_0$ , vagy az  $F_1$  frekvenciáján tüzelnek, az egyik prominencia tehát mindenképpen dominálja a tüzelési mintázatot – ilyen értelemben tehát ezek a hangzók csak a nyílt vagy legalsó nyelvállású magánhangzóktól különülnek el egyértelműen, melyek esetében az  $f_0$  és az  $F_1$  között vannak saját frekvencián tüzelő idegsejtek is (Stevens 1998).

Az előbbiektől feltehetőleg nem független jelenségre és percepciós működésekre alapozza elméletét Chistovich és Lublinskaja (1979). Tanulmányukban pszichoakusztikai kísérletek során azt vizsgálták, hogy hogyan feleltethető meg egy két prominenciát tartalmazó akusztikus jel egy egy prominenciájú akusztikus jelnek. Ehhez a szerzők olyan (szintetizált) mintahangokat generáltak, amelyeknek két prominenciája, azaz formánsfrekvenciája volt, és az egyes stimulusokban ez a két prominencia szisztematikusan más-más távolságra helyezkedett el egymástól. A kísérletben részt vevők feladata az volt, hogy „keverjék ki” ezt a kétformános jelet egy egyformános jelből, melynek a spektrális maximumát ( $F'$ ) állíthatták. Az eredmények azt mutatták, hogy ha a mintahangban a két formáns egy kritikus távolságon belül helyezkedett el (egymáshoz képest), a jelet a tesztelők egyöntetűen egy olyan egyformános jellel képezték le, ahol az  $F'$  frekvenciaértéke a két formáns közötti, az amplitúdóval súlyozott átlagos frekvencia, azaz a „gravitációs középpont” (center of gravity) frekvenciaértékét vette fel. Ha azonban a két formáns a kritikus távolságnál messzebbre került egymástól, az  $F'$  értékét a hallgatók (a formánsok relatív amplitúdójától is függően) vagy a mintajel  $F_1$ - vagy  $F_2$ -értékére hangolták, tehát a két prominenciát külön-külön érzékelték. Így Chistovich és Lublinskaja megállapította, hogy egy kritikus távolságon belül, mely  $3\text{--}3,5$  Bark<sup>5</sup>, a spektrális prominenciák az észleletben „összegződnek” (bár matematikai értelemben inkább kiátlagolódnak) egyetlen spektrális prominenciába, a két prominencia „gravitációs középpontjába” (azaz súlyozott átlagába).

---

<sup>5</sup> A Bark a Hertz értékkel szemben nem lineáris, hanem egy érzeti transzformáción alapuló akusztikai mértékegység. Erről és más nemlineáris mértékegységekről a fejezetben később még teszünk említést.

Ugyanezt a jelenséget vizsgálták Carlson és munkatársai (1970) ötformánisos szintetizált svéd magánhangzók segítségével, melyekhez a tesztelők egy kétformánisos jel második prominenciáját (tehát F2-jét) hangolhatták. A vizsgálat kimutatta, hogy az F2' értékét a hallgatók általánosan a mintajel F2-értékére hangolták a veláris /a o u/ magánhangzók esetében (ahol az F2 értéke relatíve alacsonyabb, és távol van az F3-tól). Ugyanakkor a palatális /e y ø ʉ æ/ magánhangzók esetében a hangolt F2' értéke a mintajel F2 és F3 formánisa közti „gravitációs középpont” frekvenciaértékét vette fel, a zárt, ajakréses (/i e/) hangzók esetében pedig az F3 frekvenciaértékéhez közeli frekvenciát (vagy az /i/ esetében ennél is magasabb értéket) kapott. A vizsgálat tehát megmutatta, hogy ha az F2 és F3 formánysok távolsága kisebb volt, mint 3 Bark, ahogy a palatális magánhangzók esetében tapasztalható, a két formáns egyetlen, szélesebb sáv szélességű prominenciává kapcsolódott össze az észleletben, míg a veláris hangzók esetében, ahol az F2 és F3 távolsága nagyobb, mint a kritikus távolság, az F2-t különálló prominenciaként, pontosan azonosították a hallgatók. A szerzők percepciós tesztekkel azt is bizonyították, hogy a vizsgált svéd magánhangzók egyértelműen beazonosíthatók akkor, ha az első két formáns alapján szintetizálják őket, és arra is következtettek, hogy a magasabb formánsfrekvenciák feltehetően inkább az egyéni hangszínezet, és nem a magánhangzóhangszín meghatározása szempontjából fontosak. Ugyanakkor a vizsgálat fontos megfigyelése az is, hogy a palatális hangzók kétformánisos közelítése a futtatott percepciós tesztekben nem felelt meg olyan pontosan az ötformánisos hangzók hangszínének, mint a velárisoké, ahol az F2' az F2 értékével egyező volt (Stevens 1998 kiemelése). Carlson és munkatársai (1970) eredményei tehát arra mutatnak, hogy a „gravitációs középponti” hatás szerepet játszik a palatális és veláris magánhangzók megkülönböztetésében, azaz annak az eldöntésében, hogy a megítélt hang hol helyezkedik el az elől- és hátulképzettség dimenzióján. Haugyanis az F2 és F3 a kritikus 3-3,5 Bark távolságon belül van, a hang az észleletben palatális lesz, míg ha a két prominencia a kritikus távolságnál nagyobb távolságra van egymástól, a hangzót velárisként azonosítjuk.

Fontos azonban megjegyezni, hogy két spektrális prominencia egymáshoz mért kisebb távolsága nem csak a „gravitációs középponti” hatás szempontjából fontos. Ha ugyanis két rezonancia (pl. az F1 és F2) egymáshoz közelebb esik, az mindkét rezonancia amplitúdónövekedését is okozza (vö. Fant 1960). Ezzel együtt viszont az is igaz, hogy egy tö-



lük távolabb eső harmadik rezonanciát (pl. F3) relatíve alacsonyabb amplitúdó jellemez, mint abban az esetben, amikor a rezonanciák távolsága kiegyenlített(ebb), tehát az F2 közelebb van az F3-hoz. Ha tehát a két rezonancia, azaz az F1 és F2 közel van egymáshoz egyúttal azt is jelenti, hogy az F3 amplitúdója alacsonyabb. Ez a megfigyelés Stevens (1998) értelmezésében megerősíti azt az elképzelést, hogy a magánhangzó-minőség szempontjából inkább az első két formáns, és kevésbé az F3 lehet döntő szerepű.

Traunmüller (1981) a „gravitációs középponti” hatást az f0 és az F1 viszonylatában vizsgálta, mégpedig a magánhangzók zártságészleletének tekintetében. Eredményei szerint a kritikus távolság hatása az f0 és az F1 viszonyában is fontos, mégpedig azért, mert a felső nyelvvállású, alacsony F1-értékkel rendelkező hangok spektrumát az észleletben egyetlen (az F1 értékének megfelelő) frekvenciakomponens dominálja, míg a nyíltabb ejtésű, magasabb F1-értékkel rendelkező hangokét ugyanezen a tartományon két prominencia, melyek az f0 és az F1 frekvenciáinak felelnek meg. Követekésképpen Traunmüller konklúziója az, hogy az f0 és F1 között lévő tonotopikus távolság<sup>6</sup> egyértelműen befolyásolja a zártságészleletet<sup>7</sup>. Syrdal és Gopal később megmutatja, hogy a magas nyelvvállású magánhangzók estén az f0 és az F1 általánosan 3 Barknál kisebb távolságra esik egymástól a produkcióban, függetlenül attól, hogy azok milyen frekvenciájúak, tehát az f0 és F1 távolsága kisebb 3-3,5 Barknál mind a gyermekek, mind a felnőtt férfiak, mind pedig a felnőtt nők ejtésében (Syrdal–Gopal 1986). Ez Traunmüller vizsgálataival és Traunmüllernek az f0 és F1 közti, a zártságészlelet szempontjából kritikus távolság szerepéről alkotott elméletével összhangban van, és alátámasztja azt.

Ugyanakkor a zártság észleletéhez további akusztikai jellemzők kapcsolhatók. Ugyanis már a magánhangzók akusztikai szerkezetét először szisztematikusan megvizsgáló Peterson és Barney (1952) is megfigyelte, hogy (felolvasásban, izolált szavakban) a magasabb nyelvvállásfokon képzett, zártabb hangzók f0-értéke is magasabb, mint az alacsonyabb nyelvvállásfokon,

---

<sup>6</sup> A tonotópiás szerveződés azt jelenti, hogy a különböző frekvenciákra érzékeny területek különböző területeken lokalizálódnak, ráadásul az alaphártya esetében ez a szerveződés igen szisztematikus is: a szomszédos frekvenciákra érzékeny területek egymás mellett találhatók (vö. Kállai et al. 2008).

<sup>7</sup> Traunmüller szándékosan a „zártság” terminust használja a felső vagy alsó nyelvvállású hangzók helyett, mert terminológiájával nem a nyelvvállásfokra, hanem az állkapocsnyitás nagyságára kíván utalni, ami vélekedése és az által hivatkozott szakirodalmi tételek szerint jól korrelál a percpeccios „zártsággal”.

nyíltabban képzett hangzóké. Az  $f_0$  ilyen módon magánhangzófüggő, intrinzikusnak feltételezett értéke tehát erősítheti a zártságészleletet az  $f_0$  és  $F_1$  különbségének csökkentése révén.

Az itt említett elméletekben fontos vonás, hogy a magánhangzókat, azok kategóriáit nem a lineáris vagy abszolút fizikai értékeken alapuló adatok alapján, hanem a lineáris adatok érzeti transzformáltjain alapulva vizsgálják – így tehát a magánhangzók variábilis megvalósulásainak kérdését is a percepció felől közelítik meg. A korábban említett „normalizálási” probléma egyik fő kérdése, hogy hogyan dolgozódik fel a percepciós rendszerben az, hogy a magánhangzók változatosan és egymással nagy átfedésben jelennek meg akkor, ha azokat a lineáris Hertz-értékek szerinti  $F_1 \times F_2$  magánhangzótérben kezeljük és ábrázoljuk. Ezt a kérdést azonban valamely nemlineáris, frekvenciavetemítésen alapuló, érzetileg transzformált skála alkalmazása (azaz például a logaritmikus vagy a Mel vagy Bark értékekbe történő transzformáció)<sup>8</sup> gyakorlatilag teljesen kikerüli, ugyanis a lineáris adatok alapján felvett magánhangzótérben tapasztalható átfedések szinte teljesen megszűnnek akkor, ha a magánhangzótér ábrázolásához ilyen érzeti skálákat, tehát egy másfajta kétdimenziós projekciót alkalmazunk (vö. Neary 1989).

A beszédpercepció szakirodalmának egyik alapvetően elfogadott nézete szerint a koartikulációs hangátmenetekben kódolt, a magánhangzók képzésére is vonatkozó akusztikai információknak köszönhetően a magánhangzók nagyobb arányban ismerhetők fel, ha a hallgatók mássalhangzó-környezetben ejtve hallgatják őket. Strange és munkatársai (1976) /p–p/ hangkörnyezetbe ágyazott magánhangzók vizsgálatakor azt találták, hogy míg a mássalhangzó-környezetben ejtett magánhangzók azonosítása 80-90% körül mozgott, az izolált ejtésű hangzókat 30-40%-ban tévesztették más hangokra a hallgatók, annak ellenére, hogy formánsértékeiket tekintve az izolált ejtésű hangzók a mássalhangzó-környezetbe ágyazott megvalósulásokhoz képest túlartikuláltak voltak, tehát egymástól inkább eltérőek, kontrasztosabbak.

---

<sup>8</sup> Míg a logaritmikus transzformáció során a frekvenciaértékeket zenei félhangokká transzformáljuk, a Bark skála az auditív maszkoláson alapszik, és az ahhoz kapcsolódó hallási eseményeket modellező szűrőket használ a transzformációhoz. Ez utóbbit, tehát a Bark szűrősort használó transzformációt pszichoakusztikai kísérletekkel dolgozták ki. A logaritmikus és a Bark szűrősorton alapuló transzformáción közti különbség valójában nem nagy, és főképp az 500 Hz alatti tartományt érinti. Ez a különbség abban áll, hogy 500 Hz alatt a Bark skála a logaritmikus skálánál lineárisabban függ össze a Hertz értékekkel, ez pedig realitásabban képezi le a hallási érzékenységet (Neary 1989).

Ugyanezen tanulmány második kísérletéből az is kiderül, hogy a mássalhangzók pozitív hatása még akkor is kimutatható, ha a mássalhangzók minősége nem megjósolható (tehát bemutatásonként változik). Jenkins és munkatársai (1983) később azt is kimutatták, hogy a formánsátmenetekben rejlő akusztikai információ még a magánhangzó mediális (tiszta fázisának tekinthető) 50-60%-ának törlése mellett is nagyon magas arányban megőrzi a magánhangzó-minőségre vonatkozó információt – ugyanis nem tapasztaltak különbséget a közepén törölt szótagra és a teljes szótagra mutatkozott azonosítási arányok között. A magánhangzó és az őt követő vagy megelőző mássalhangzó közti formánsátmenetek mint dinamikus akusztikai kulcsok, azaz a koartikuláció akusztikai vetülete tehát úgy tűnik, nagyon fontos akusztikai kulcsokat jelent a magánhangzók azonosításában.

Neary (1989) az akusztikai információ négy típusát különíti el mint az észlelésben (legalábbis az angol magánhangzók észlelésében) fontos, de változó súlyú kulcsokat. Az első a) típusba a statikus tulajdonságok tartoznak, mint például a magánhangzó középső, nagyjából állandó szeletének (ezt magyarul tiszta fázisnak is nevezik, l. pl. Gósy 2004) formánsfrekvenciái és alapfrekvenciája. A második b) típusba Neary a dinamikus tulajdonságokat sorolja – ilyenek a magánhangzóban rejlő inherens spektrális változások és a mássalhangzó-környezet hatására megjelenő formánsátmenetek, melyekkel kapcsolatos kísérletes eredményekről az imént tettünk említést. A harmadik c) csoport az intrinzikus, azaz a szegmentumon belüli frekvenciarelációkból fakadó tulajdonságok csoportja, ide különösen a formánsok egymáshoz és az alapfrekvenciához viszonyított relációja tartozik. A negyedik d) csoportba pedig az extrinzikus, azaz a transzsegmentális relációs tulajdonságok tartoznak, mint az adott magánhangzónak az adott beszélő többi magánhangzójához viszonyított relatív időtartama, valamint a magánhangzó a többi magánhangzó formánsértékeihez viszonyított relatív frekvenciaszerkezete – ezek az információk ugyanis viszonyítási pontokként segítik a percepció „kalibrálását” a beszélő egyéni sajátosságaiból adódó hatások eliminálása (azaz a beszélőnormalizálás) révén (l. pl. még Ladefoged–Broadbent 1957).

Mindabból, amit a beszédhangok és különösen a magánhangzók észleléséről eddig tudunk, és amely tudásnak egy szeletét a jelen alfejezetben is igyekeztünk – a teljesség igénye nélkül – körüljárni, fontosnak tartjuk kiemelten hangsúlyozni a következőket. A sok kísérletes munka

eredményeképpen ma már sejthető, hogy melyek azok az információk, amelyek elegendőek a magánhangzó-azonosításhoz, vagy melyek azok, amelyeket a percepció működése hasznosítani tudnak. A normál módú beszédben minden bizonnyal kiemelt szerepe van az F1 és F2 formánsoknak (valamint, főként az F2-vel való auditív interakciója miatt, az F3 formánsnak a palatálisok esetében), továbbá az f<sub>0</sub>-nak, azaz az alapprofrekvenciának. Úgy tűnik ugyanis, hogy ezek az akusztikai információk és viszonyaik elegendő információt hordoznak a magánhangzók minőségének beazonosításához. Azt azonban, hogy ezekből az információkból pontosan mi elengedhetetlen a beszédpercepció számára, hogy mi történik akkor, ha az alapprofrekvencia emelésének következtében ezek az információk nem elérhetők (vö. az előző fejezetben említett alul-mintavételezettség fennállása esetén), vagy hogy ezeket az információkat pontosan milyen módon hasznosítja az emberi percepció pontosan, nem tudjuk biztosan.

#### **1.1.2.3. A magyar magánhangzók azonosításának kísérletes eredményei**

A magyar magánhangzók azonosítását, illetve a frekvenciaszerkezet percepció hatását a magyar magánhangzók minőségének felismerésében Gósy Mária (1989) vizsgálta behatóan egy három fázisból álló kísérletsorozatban.

Az első fázisban Gósy (1989) a 14 magyar magánhangzót egy férfi beszélő bemondásában rögzítve, izolált ejtésben azonosíttatta a hallgatókkal, mégpedig úgy, hogy a magánhangzókat kilencféle különböző felső határértékre beállított aluláteresztő szűrővel szűrte. A kísérlet több tesztből állt, melyek során a szűrő felső határának frekvenciaértékét fokozatosan növelték, így az áteresztett tartomány, tehát az áteresztett akusztikai információ mennyisége is fokozatosan nőtt a kísérlet előrehaladtával. Az eredmények szerint egyes hangzók (elsősorban az /ɒ/) esetében már az F1 megjelenése előtt (tehát akkor, amikor az F1 frekvenciaértéke magasabb volt, mint a szűrő felső határának értéke) lehetséges volt a magánhangzók helyes azonosítása, továbbá némely más hangzók esetében (/o o:/) bár szükséges volt az F1, de az F2 megjelenése előtt (azaz az F2 alatti frekvenciasáv alapján) is biztosítottak tűnt a magasabb arányú helyes azonosítás. Ugyanakkor az eredmények azt is mutatják, hogy egyes hangok esetében nem hogy javított, hanem inkább még rontott is az azonosítási arányokon az egyre részletesebb akusztikai információ: míg az /i/ a 390 Hz felső határértékkel rendelkező frekvenciasáv

(azaz az F1-et tartalmazó spektrumú hangrészlettel) 80%-ban azonosíthatónak bizonyult, az egyre növelt frekvenciájú felső határ egyre alacsonyabb azonosítási arányokhoz vezetett. Ezekből az adatokból a szerző azt a megállapítást teszi, hogy az F1-et tartalmazó frekvenciasáv valamilyen formában tartalmazza az F2 információt is, valamint azt, hogy az F1 és F2 közé eső tartomány nem erősített frekvenciakomponensei is fontos szerepet játszanak a magánhangzó-minőségek felismerésében. Bizonyos értelemben tehát a szerző hasonló következtetésre jut, mint néhány korábban idézett vizsgálat lefolytatói. Surlovicz és Goldstein (1983), valamint Delgutte és Kiang (1984) ugyanis vizsgálataikban szintén hangsúlyozzák az F1 és F2 közti frekvenciatartomány fontosságát. Emlékeztetőül, az idézett kutatók megállapítása szerint a hallóidegek tüzelési mintázatai alapján az egymáshoz adott (3-3,5 Barknál kisebb) távolságra elhelyezkedő F1 és F2 rezonanciák „összeolvadnak”, az egymástól ennél nagyobb távolságra elhelyezkedő rezonanciák viszont elkülönülnek az észleletben – ilyen értelemben tehát az F1 és F2 közti frekvenciatartomány megléte és kiterjedése elválasztja egymástól a magánhangzók egyes csoportjait. A tévesztések tendenciái a legalacsonyabb felső határértékkel szűrt hangmintákon a szerző összefoglalása szerint elsősorban zárt, vagy felső nyelvvállású, hátul képzett hangok (téves) felismerését jelentették. Ez a megfigyelés Traunmüller (1981) eredményeivel hozható összefüggésbe. Traunmüller (1981) ugyanis azt találta, hogy a felső nyelvvállású hangzók észleleti lenyomatában az  $f_0$  és F1 tartományán mindösszesen egy észleleti prominencia van jelen (míg a nyíltabb hangzók esetében kettő, mégpedig az  $f_0$  és az F1 frekvenciáján), nem meglepő tehát, ha egy olyan jelet, amelyben formánsinformáció híján mindösszesen az alaphang jelenik meg prominenciaként, inkább zártabb hangzóként azonosítanak az észlelők.

Gósy (1989) a kísérlet második fázisában sáváteresztő szűrőket alkalmazott az izoláltan ejtett magánhangzók akusztikai módosítására, melynek felső határa minden esetben 2700 Hz volt, alsó határát azonban a szerző fokozatosan állította 270 Hz-től indulva 2200 Hz-ig (így a szűrés után létrejött hangzók sáv szélessége a fokozatokkal egyre csökkent). A kísérlet legfőbb tanulságaként Gósy (1989) azt a következtetést vonta le, hogy a sávszűrt magánhangzók azonosítása a vizsgált sávokon belül végig pontos maradt, azaz egyes hangzók azonosítása még akkor is viszonylag magas arányú maradt, amikor az F1 és F2 formáns gyakorlatilag teljesen ki volt szűrve a jelből, és csak az F3 esett a szűrő által áteresztett tartományba (bár ez alól a

tendencia alól azért egyes hangzók kivételt képeztek). A szerző megjegyzi, hogy a legmagasabb alsó határértékkel szűrt tartományokban inkább az elől képzett hangzók azonosítása volt pontos, még akkor is, ha azok F2-je a kiszűrt tartományba esett, azaz ha az áteresztett tartományban csak az F3 és nem erősített frekvenciahelyek maradtak. Ebből a szerző az F3 kiemelt jelentőségére következtetett: összegzése szerint az F3 jellemző és meghatározó a magánhangzó-minőség szempontjából, mégis inkább csak másodlagos felismerési kulcsnak tekinthető, hiszen az F1 és F2 jelenlétében pusztán az F3 manipulálása nem befolyásolja a fonémadöntést, azok hiányában azonban az F3 jelentősége megnő. Ismét csak korábbi vizsgálatokkal párhuzamot vonva megemlítendő, hogy Gósy (1989) konklúziója összhangban van Carlson és munkatársainak (1970) eredményeivel. Azok szerint ugyanis a palatálisoknál az F2 és F3 percepció közelsége miatt az F2 és F3 helyén valójában nem két, hanem csak egy, frekvenciájában a két formáns közé eső érzeti prominencia jelenik meg – ilyen értelemben tehát az F3 az F2-vel „összefonódva” képez egyetlen akusztikai kulcsot, nem tekinthető tehát másodlagos kulcsnak a palatálisok azonosításában.

A vizsgálat harmadik fázisában Gósy (1989) felüláteresztő szűrő segítségével manipulálta a beszédhangokat, mégpedig úgy, hogy a szűrő alsó határát módosította 1000 Hz és 3300 Hz között. Az eredmények szerint a legmagasabb alsó határértékkel szűrt hangminták közül a legmagasabb arányban az /i i:/ és /y y:/ magánhangzókat azonosították a hallgatók, melyek második formánsa egyébként a legmagasabb volt a kísérletben. Az /i i:/ és /y y:/ egymástól való elkülönítését, azaz az ajakkerekítésre vonatkozó információ akusztikai kulcsolását a szerző konklúziói szerint két tényezőben látja biztosítottnak. Az egyik tényező a részösszetevők eltérő intenzitása, a másik pedig az az ismert összefüggés, hogy az ajakréses hangzók intenzitása általában véve nagyobb, mint az ajakkerekítéseseké (főként azért, mert az ajakkerekítésesek esetében nagyobb az ajaknyílás-felület, mely egységnyi idő alatt nagyobb levegőmennyiséget, de következésképpen kisebb teljesítménnyel ereszt át).

Egy további vizsgálatban Gósy (1989) az időtartam és a frekvencia-összetevők szerepét vizsgálta szintetizált magánhangzók azonosításában. Ebben a percepció kísérletben a résztvevőknek különböző frekvenciaszerkezetű és különböző időtartamú magánhangzókat kellett azonosítani. Az eredmények azt mutatták, hogy egyre növekvő időtartam mellett a hallgatók az

F1 értékek egyre szűkebb tartományában azonosítottak egy adott hangzóminőséget, az időtartam növekedésével ugyanis inkább nyíltabb magánhangzókat véltek felismerni (pl. /u~u:/ helyett /o~o:/-t). (Megjegyzendő, hogy ez alól a tendencia alól kivételt képeztek a labiális palatális /y~y: ø~ø:/ hangzók, valamint hogy a bemutatott adatok szerint nem fordult elő, hogy egy magánhangzó az észleletben két kategóriát „ugrott” volna.) A kísérlet tanulságát a szerző úgy foglalja össze, hogy megállapítja, a magánhangzó időtartama fontos kulcs a magánhangzóminőség felismerésében: bár azokban a helyzetekben, amikor a szintetizált magánhangzók formánsértékei a magyar magánhangzókra legjellemzőbb formánsfrekvenciák értékeit vették fel, az időtartam mint akusztikai kulcs másodlagosnak bizonyult, más esetekben az időtartam a frekvenciaszerkezettel szimultán, elsődleges kulcsnak mutatkozott.

Gósy (1989) kísérleteit később főként olyan magánhangzó-azonosítási vizsgálatok követték, melyek a magánhangzók fizikai és nyelvi időtartamának összefüggéseit, valamint a nyelvi időtartam és a frekvenciaszerkezet összefüggéseit elemezték elsősorban az elméletileg csak fizikai időtartamukban különböző felső és középső nyelvválású párok esetében (pl. Kovács 1998, Mády–Reichel 2007). Ezen vizsgálatok bemutatására a jelen értekezés keretei között azonban nincs módunk, hiszen – bár a magyar magánhangzórendszert érintő fontos kérdéseket vizsgálnak auditív szempontok szerint –, a jelen értekezésben ismertetendő kísérletek felől nézve a fonológiai oppozíció tagjainak megvalósulására és auditív oppozíciójára vonatkozó eredmények nem relevánsak. Bemutatandó kísérleteink ugyanis (az énekelt magánhangzók azonosítására fókuszálva) e párok tagjai közül minden esetben csak a hosszú magánhangzókat vizsgálják megközelítőleg egységesen hosszú időtartamban (kitartott ejtésben), mert elsősorban nem a magánhangzók észlelt nyelvi hosszúságának, hanem minőségének változásait kívántuk elemezni az éneklésben (vö. a *Kísérletek* fejezet *Anyag, módszer és kísérleti személyek* alfejezetében található leírásokkal.)

Összefoglalóan elmondhatjuk tehát, hogy a magyar magánhangzók azonosításában a nemzetközi szakirodalomban tapasztalt egyes tendenciákat már kísérletes úton is igazolták, megerősítést nyert az F1 és F2 formánsok kiemelt szerepe a hangzóminőség perceptuális kialakításában – de azt is láttuk, hogy nem feltétlenül szükséges mindkét formáns teljes sáv szélességű megléte a magasabb arányú azonosításhoz.

A jelen értekezés vizsgálatainak középpontjában egy sajátos beszédképzési móddal létrehozott magánhangzók állnak: az énekelve ejtett vagy énekelt magánhangzók. A témamegjelölés azt sugallja, hogy ez a kérdéskör elkülönül a normál beszéd módban ejtett magánhangzók percepciójának problémájától – a következő fejezet célja az, hogy az éneklésről eddig ismeretes artikulációs, akusztikai és percepció-s sajátosságok tárgyalásán keresztül megvilágítsuk, ez valóban így van. A következő fejezetben tehát az énekelt magánhangzók szakirodalmának a lehetőségek szerint alapos, de minden bizonnyal nem teljes összefoglalására törekszünk figyelmünket elsősorban az énekelt magánhangzók felismerésének kérdése szempontjából fontosabb aspektusokra fókuszálva.

## **1.2. Az éneklés és az énekelt magánhangzók fonetikai jellemzése**

Az éneklés fonetikai szempontokat érvényesítő szakirodalmának összefoglalása sem kezdődhet másként, mint az „éneklés” fogalmának pontos meghatározásával. Mivel azonban az értekezés nem kíván szólni sem zenetörténeti, sem pedig zeneelméleti kérdésekről, a fogalom meghatározásában sem törekszünk (és nem is törekedhetünk) a zenetudomány felől nézve maradéktalanul pontos meghatározásra. E „definíció” célja sokkal inkább az egyértelmű és pontos beazonosíthatóság kell, hogy legyen. Ennek fényében úgy fogalmazhatunk, éneklésen a jelen értekezésben a tradicionális, vagy nyugati típusú operai műfajt értjük, mitöbb, a jelen értekezésben e műfajon belül is elsőként a magas alapfrekvenciájú szopránéneklésre fókuszálunk.

Az opera hagyományosan igen sajátos műfaj – bizonyos értelemben mind az igen korai, mind pedig a kortárs énekműfajokkal szemben áll. A mai zenei műfajokhoz hasonlóan ugyanis nagy hangerőt igényel – hiszen az operai szólóénekeseknek egész zenekarokat kell „túlénekelnie” –, mégis énekesei saját beszédképző szerveiken kívül hagyományosan nem használnak semmilyen (külső) hangerősítést. Az olasz opera megjelenésekor a reneszánsz világ polifonikus zenéjében különleges újdonságot jelentett, hiszen alapvetően szólisztikus műfaj. Európai elterjedésekor, a 17. század környékén a műfajt még elsősorban az igen lágy hangjukról, finom díszítésekkel jellemezhető stílusokról ismert kasztrált énekesek (*castrati*) dominálták. Az



ekkoriban lezajló zenei fejlődés, a zenei műfajok változása azonban a használt hangszerek felépítésének változásaival is együtt járt, az egyre jobb rezonátorképességekkel megépített hangszerek hangereje egyre növekedhetett, így a halkabb szólisták ideje hamar leáldozott. A használt hangszerek cseréje és a hangszerek felépítésének megváltozása mellett ráadásul a kísérőzenekar mérete is gyarapodott, együtt a zenei eseményeknek helyt adó terek növekedésével. Mindennek eredményeképpen tehát a 19. századra nem csak az énekhang minőségének, de „mennyeségének” is módosulnia kellett, hogy az egyre gyarapodó auditív maszkoló (azaz elfedő, a hallhatóságot akadályozó) jelenségeket leküzdje. Így alakultak ki egyes új hangminőségek (a koloratúr és drámai szoprán), alakultak át az egyes szerepek (egyre nagyobb szerepet kaptak a magas alapfrekvenciákon éneklő énekesek, a tenorok, baritonok vagy a mezzoszopránok), elkezdett egyre feljebb és feljebb terjeszkedni a tenor és szoprán hangterjedelem – így jött létre tehát (nagyjából az 1800-as évek végére) a ma is ismert operai hangzás és énektechnika. Az énekhang minősége vagy az énekhang hangszínezete azonban nem pontosan definiált, a technika örökítésében alapvetően a mai napig elsősorban egy-egy tanár reprezentálja a stílusbeli és hangzásbeli kívánalmakat (Mitchell 2005).

A ma ismert operaéneklés hagyománya tehát több száz éves, az operai énekhang fontos kritériuma a nagy hangerő és a lehetőség szerint a teljes alapfrekvencia-tartományon homogén hangszínezet. Bár ma már rendelkezésre áll a hangerősség növelésére számos (elektro)technikai megoldás, a szigorúan vett operai műfaj (szemben például a musical vagy operett műfajokkal) nem él ezekkel a lehetőségekkel – a hagyományos kritériumoknak megfelelő hangszínű és hangerejű hangadás kulcsa pusztán az énekes rezonátüregének jobb, hatékonyabb kihasználásában rejlik.

A jelen értekezés témája a magas alapfrekvenciájú szopránéneklés. Ennek oka, hogy vizsgálatok tanúsága szerint a magasabb frekvenciák énekesei – de elsősorban a szopránok – az alacsonyabb hangosztályok (basszus, bariton vagy alt) énekeseihez képest egészen más jellegű akusztikai problémák megoldására kényszerülnek, melyeket (értelemszerűen) erősen eltérő technikával, eltérő stratégiákkal oldanak meg. Ezen túlmenően a jelen vizsgálatok elsősorban a magánhangzókra fókuszálnak, hiszen az énekes bizonyos értelemben csak e hangok segítségével „énekel” – mivel a magánhangzók képzésekor a levegő áramlása akadálytalan, a ma-

gánhangzós részek nem tartalmaznak zörejes elemeket, csak a magánhangzók képzését általánosan jellemző zöngét, így ezek a hangzók a legalkalmasabbak az alaphangélmény előidézésére, azaz a dallam kifejezésére, így alapvetően ezeken a hangzók alapszik a *bel canto* néven is ismeretes operai stílus.

A magas alapfrekvenciájú éneklés problémásnak tekinthető mind a hangerő, mind a hangszínezet, mind pedig a hangszín, azaz az egyes magánhangzó-minőségek szempontjából. Érdekes módon azonban mindezek – úgy tűnik – ugyanazon fonetikai problémára vezethetők vissza, mégpedig arra, hogy a szoprán énekesek gyakran kényszerülnek olyan magas alapfrekvenciájú hangok kiéneklésére, melyek a beszédhang minősége szempontjából kiemelten fontosnak tartott első formáns frekvenciájánál magasabbak. Az F1 „túllövése”, tehát a magánhangzók változatlan ejtése magasabb alapfrekvenciákon egyrészt hangerőbeli csökkenést okozna, másrészt pedig olyan erős hangszínezetbeli váltást is, mely ebben a műfajban nem kívánatos (vö. pl. Sundberg 1987). Ráadásul – mindabból, amit a magánhangzók azonosításáról eddig tudunk – feltehető, hogy az F1 eltűnése a magánhangzóspektrumból az ejteni szándékozott magánhangzó felismerhetőségét is erősen csökkentené. Amint azonban egy operaária meghallgatása során ezt magunk is tapasztalhatjuk, képzett énekesek esetében ezek a problémák nem, vagy legalábbis csak részben állnak fenn – professzionális énekesek esetében ugyanis alapvetően nem jellemzők a jelentős hangszínezetbeli váltások, és nem is nyomja el az énekes hangját a kíséret. Ugyanakkor a magánhangzók néha bizonytalan minőségét és az énekelt szövegek értési nehézségeit feltehetően mindannyian tapasztaltuk már. Kérdés tehát, hogy mit tesz az énekes annak érdekében, hogy hangja homogén hangszínezetű és nagy hangerőségű maradjon magasabb alapfrekvenciákon is, és mi történik ekkor az ejteni szándékozott magánhangzók minőségével.

A magas alapfrekvencián énekelt magánhangzókat vizsgáló szakirodalomban ma már szinte közhelynek számít az a feltételezés, mely szerint azokban a helyzetekben, amikor az F1 értéke alacsonyabb lenne, mint az f0 frekvenciaértéke (rövid jelölése:  $F1 < f0$ ), az énekes az F1 értéket a megemelt f0 értékére vagy valamivel afölé hangolja (vö. pl. Sundberg 1987).<sup>9</sup>

---

<sup>9</sup> Itt jegyezzük meg, hogy a jelen értekezésben a formánshangolási stratégiákat a nemzetközi szakirodalom (l. pl. Joliveau et al. 2004, Garnier et al. 2010) egyik elfogadott konvenciójának megfelelően kettősponttal jelöljük

Sundberg (1979 és 1987) szerint (aki a szoprán énekesek alapfrekvenciafüggő produkciós tendenciáit elsőként tanulmányozta mélyrehatóan, és akinek vizsgálatairól a következőkben még sok szó lesz) ez az  $F1 : f_0$  hangolásról szóló feltételezés akár empirikus bizonyítékok nélkül is igen stabil lábakon áll. Állítása szerint ugyanis már pusztán az  $F1$ -nek az  $f_0$ -ra hangolásával, azaz a megszólaló hang alapfrekvenciájának felerősítésével 20 vagy akár 30 dB hangnyomásszint-növekedés (azaz hangerő-növekedés) is elérhető lehet. A formánshangolás, azaz a toldalékcső rezonanciatevékenységének e jobb kihasználása ráadásul úgy idézi elő a hangerő-növekedést, hogy ahhoz semmilyen további vokális erőfeszítés (pl. a szubglottális nyomás növelése, vö. kiabálás) nem szükséges (l. még Titze 1994 és 2004). Ez a felvetés azonban empirikus úton igen nehezen igazolható, hiszen a magas alapfrekvenciákon ejtett magánhangzók akusztikai lenyomatában (mint azt már az előző fejezetben említettük) az alulmintavételeztség jelensége miatt a hagyományos, Fourier-transzformáción alapuló elemzések segítségével nem határozhatók meg megbízhatóan az egyes formánsok középértékei, így az  $F1$  frekvenciája sem. Ennek eredményeként tehát Sundberg első vizsgálatait – és az  $F1 : f_0$  hangolás első felvetését – számos további vizsgálat követte, melyek vagy közvetett, artikulációs adatokból derivált akusztikai adatok alapján, vagy közvetlen, valamely újszerű vagy szokatlan mérési módszertan alkalmazásával nyert akusztikai adatokból igyekeztek meghatározni az énekelt magánhangzók ejtésekor jellemző rezonanciafrekvenciákat. E vizsgálatok fő kérdése, hogy milyen módokon képes a magas alapfrekvencián éneklő szoprán énekes rezonátorüregeinek jobb kihasználására, így nagyobb hangerő és szép, homogén hangszínezet létrehozására, továbbá – egyes kísérletekben – az a kérdés is felmerül, hogy miként hatnak ezek a módosítók az egyes magánhangzó-minőségek függvényében.

A formánshangolás megléte ugyanakkor feltételezhetően nemcsak a hangerő növekedését idézi elő, hanem az egyes magánhangzó-minőségek produkciójára, az egyes minőségek közti különbségek fenntartására is hatással van. Az  $F1 : f_0$  hangolás eredményeként ugyanis a magasabb alapfrekvenciákon az egyes magánhangzók akusztikai szerkezete (az egységes  $f_0$  és  $F1$  érték miatt) egymáshoz a beszédben tapasztalhatónál jóval hasonlóbb lesz, ami valószínűsíthe-

---

olyan módon, hogy a kettőspont bal oldalán a hangolt formánst, a jobb oldalán pedig a hangolási célt tüntetjük fel, például:  $F1 : f_0$ , melynek jelentése „az  $F1$ -nek az  $f_0$ -ra hangolása”.

tő módon a perceptuális különbségek csökkenését is maga után vonja. Ebből a feltételezésből következően az énekelt magánhangzók vizsgálatának másik nagy területe a magánhangzók percepciójának kutatása – eddig ezen belül is elsősorban az azonosíthatóság határait és segítő körülményeit kutatták. Mivel az énekelt és a beszélve ejtett magánhangzók közti különbségek jóval nagyobbak is lehetnek, mint a beszédben egyébként is tapasztalható (beszélőn belüli és beszélők közötti) variabilitás<sup>10</sup>, az énekelt magánhangzók észlelésének vizsgálata egyúttal a beszédpercepció működéseiről alkotott képünket is tovább árnyalhatja – így az észlelés rugalmassága vagy akár korlátai is jobban megfigyelhetők. Nem mellesleg az énekelt magánhangzók produkálhatóságának és azonosíthatóságának feltételeiről és korlátairól szóló ismeretek a gyakorlatban is hasznosíthatók lehetnek: felhasználhatók az énektechnika tanításában, az operaszövegek fordításában, vagy figyelembe vehetők akkor, mikor egy-egy operaelőadás feliratozásának szükségességét latolgatják az abban illetékesek. Mégis, érdekes módon, az akusztikai és artikulációs vizsgálatokhoz képest az énekelt magánhangzók észlelésének vizsgálata jóval alulreprezentáltabb a szakirodalomban, és az eddigi eredmények sok kérdést megválaszolatlanul hagynak.

A jelen fejezetben előbb az éneklés artikulációs és akusztikai sajátosságait vizsgáló kísérleteket vesszük sorra, melyek főként egy-egy idegen nyelv (a svéd vagy az angol) hangzókészletét elemezték, több esetben pedig csak egy-egy magánhangzót. Az összefoglalás célja, hogy képet adjon az éneklés specifikus artikulációs jellemzőiről, valamint hogy bemutassa az eddigi eredményeket, melyek a szoprán hangosztályú énekesek formánsstratégiáira vonatkoznak. A bemutatásban előbb a gégeszintű működéseiről, valamint a gégeműködés és az artikuláció összefüggéseiről lesz szó, majd az akusztikumra (azaz a formánsstratégiákra) vonatkoztatott, közvetett artikulációs és közvetlen akusztikai bizonyítékokat, eredményeket ismertetjük.

---

<sup>10</sup> Mind a beszélőn belüli, mind a beszélők közötti változatosság, annak okai és megvalósulásai számos vizsgálat tárgyát képezték, ráadásul e vizsgálatok száma egyre bővül. Ezekből itt csak párat említünk. A beszélőn belüli változatosság egyik aspektusa az életkorral összefüggő változások sora, melyek vizsgálata longitudinális (l. pl. Lee et al. 1999, Voreprian–Kent 2007) vagy keresztmetszeti (l. pl. Bóna–Imre 2010, Bóna 2012, Bóna 2014, Gósy–Bóna 2014) lehet. (Előbbi esetben az egyes kísérleti személyek beszédének változását követik a kutatók, utóbbi esetben az egyes életkori csoportokat nem ugyanazon beszélőkkel reprezentálják.) Egy másik aspektus lehet az ugyanazon személyek beszédének a beszédstílusok mentén tapasztalható eltérése (l. pl. Markó et al. 2010). A beszélők közötti változatosság anatómiai, attitűdinális, szociális vagy kulturális jellemzőkből, különbségekből is fakadhat. Ezek vizsgálatára példaként itt csak Gósy (2004), valamint Gósy és Siptár (2013) munkáit említjük. E szerzők ugyanis kiemelt hangsúlyt fektetnek a fent említett két témakör vizsgálatára.

Ezután áttérünk az énekelt magánhangzók azonosítását vizsgáló kísérletek eredményeinek ismertetésére, melyek az énekelt magánhangzók sajátos artikulációs és akusztikai tulajdonságainak percepció vetületét vizsgálják.

A fejezet zárásaképp az eddigi nemzetközi eredmények összefoglalásaként kiemeljük a máig vitásnak tűnő vagy nyitva maradt kérdéseket, megfogalmazzuk az éneklésnek mint alternatív beszédképzési módnak a magyar magánhangzókra vonatkoztatva relevánsnak tűnő (produkciós és percepció területet egyaránt érintő) vizsgálati kérdéseit, és közöljük a jelen értekezés vizsgálatainak alapját képező hipotéziseket.

### **1.2.1. Az énekelt magánhangzók artikulációs vizsgálata**

A következőkben elsőként tehát az énekelt magánhangzók korábbi artikulációs vizsgálatait mutatjuk be. Elválasztva kezelünk itt azonban két kérdést, és ennek megfelelően az artikulációs vizsgálatokat is két részre bontjuk. Külön tárgyaljuk azokat a vizsgálatokat, melyek célja elsősorban nem az akusztikumra, az  $F1 : f_0$  formánshangolásra vagy más formánsstratégiákra vonatkozó indirekt bizonyítékok keresése volt (ennek szenteljük a jelen, azaz a *1.1.1 alfejezetet*), valamint azokat, melyek artikulációs adataikkal elsősorban a szoprán énektechnikában rejlő sajátos rezonanciastatégiákat kívánták közvetett módon elemezni (ezeket az *1.1.2 alfejezetben*, az akusztikai sajátosságok alatt tárgyaljuk).

#### **1.2.1.1. A vertikális és horizontális gégehelyzet az éneklésben**

Már az első tudományos érdeklődő énektechnikai szakkönyvek említést tesznek az éneklés során alkalmazandó megfelelő függőleges gégehelyzetről. Ennek az irányzatnak úttörője Manuel Garcia (1911), aki egy az addigiakhoz képest gyökeresen új tudományos alapokon nyugvó tanításmódszertan kidolgozását tűzte ki céljául. Garcia egy kezdetleges laringoszkóp alkalmazásával képes volt megfigyeléseket tenni az éneklés közben tapasztalható fonációra és a függőleges gégehelyzetre – munkássága nyomán tehát már az 1840-es évektől terjedni kezdett az a nézet, mely szerint a jó énektechnikával képzett hang produkciója során a gége folyamatosan alacsony helyzetben van. Garcia nevéhez fűződik egyébként gyakorlatilag mindannak a hagyománya, hogy a gégehelyzetre vonatkozó vagy ehhez hasonló más anatómiai, fiziológiai

megfigyelések egyáltalán bekerültek a tanítás-módszertani és énektechnikai szakkönyvekbe (vö. Mitchell 2005).

A tudományosabb, állításait anatómiai alapokra is helyező tanítás-módszertani irodalom alapján úgy láthatjuk, a mai napig tartja magát az a nézet, mely szerint a magas gégehelyzetben keltett fonáció az éneklés során károsítja a hangszalagokat, így kerülendő (vö. pl. Vennard 1967, idézi Pabst–Sundberg 1992). Azt azonban már régóta tudjuk, hogy a fonáció nem független az artikulációtól: mind a gégehelyzetre, mind a hangszalagrezgésre hatással van az artikulációs szervek, az artikulátorokat mozgató izmok tevékenysége (vö. pl. Honda 1983, Hoole–Kroos 1998). Kérdés tehát, hogy az énekesek valóban képesek-e állandó (ráadásul alacsony) pozícióban tartani a gégejükét alaphfrekvenciától és magánhangzó-minőségtől függetlenül – és kérdés az is, hogy valóban célja-e ez az énekesnek, melyre a tanítás során őt trenírozni kell.

A gégehelyzetet az éneklés során modern eszközökkel, különféle képalkotó eljárásokkal vizsgáló szakirodalom alapján azt látjuk, ez a feltételezés nem bizonyítható maradéktalanul, sőt a vizsgálatok eredményei alapján sok esetben éppen e hagyományos feltételezés cáfolata fogalmazható meg, hiszen a kutatások sok esetben az alaphfrekvenciától (valamilyen módon) függő gégehelyzetet igazolták az énekesek énekprodukcójában.

Az énekelt magánhangzók során tapasztalható vertikális gégehelyzetet (vertical larynx position, VLP) először Johansson és munkatársai (1982) vizsgálták röntgenográfia segítségével. A vizsgálatban a szerzők a magánhangzótér artikulációs és akusztikai értelemben is szélső értékeit képező /a i u/ megvalósítását elemezték egy szoprán és egy alt énekesnél, a szoprán esetében 230 Hz, 470 Hz és 950 Hz alaphfrekvenciákon, míg az alt esetében 150 Hz, 300 Hz és 600 Hz alaphfrekvenciákon, laterális röntgenfelvételek segítségével. (Bár az elemzésekkor a gége függőleges pozícióján kívül a kutatók az ajak- és állkapocsnyitás szögét is mérték, valamint a toldalékcső keresztmetszetének kiszámítása után a toldalékcsőre az egyes artikulációs konfigurációkban jellemző rezonanciafrekvenciákat is megbecsülték, a vizsgálat ezen részeivel itt nem foglalkozunk – azokról mint a formánstratégiákhoz kapcsolódó eredményekről a következő alfejezetben lesz szó.) Egyetlen szoprán énekes résztvevő eredményei alapján Johansson és munkatársai (1982) megállapították, hogy az elvárásokkal ellentétes módon a

gége függőleges helyzete nem volt állandó az alapfrekvencia függvényében (ahogyan azt az énektechnikai szakkönyvek alapján feltételezhetnénk), valamint nem igazolódott az sem, hogy a gége a nyugalmi helyzethez képest elsősorban alacsony pozícióban helyezkedett volna el. Az eredmények szerint ugyanis az alapfrekvencia emelésével a gége megemelkedett, mégpedig az ejtett magánhangzó minőségétől nagyjából függetlenül – ugyanakkor az egyes magánhangzók esetében némileg eltérő gégepozíciót detektáltak a szerzők: a legalacsonyabb VLP-értékeket az /u/ esetében mérték, míg az /a/ és /i/ esetében ennél valamivel magasabbat. Mivel a kontrollként vizsgált, valamivel alacsonyabb alapfrekvenciákon éneklő alt esetében nem látszott korreláció a vízszintes gégehelyzet és az alapfrekvencia között, a szerzők végeredményben arra következtethettek, hogy a gége emelkedését inkább a magasabb alapfrekvenciák idézik elő.

Később Pabst és Sundberg (1992) egy többcsatornás EGG berendezéssel (tracking multi-channel electroglottograph, TMEGG) terjesztette ki a fenti kísérletet két professzionális szoprán, egy mezzoszoprán és további négy professzionális, valamint egy amatőr bariton énekes résztvevő vizsgálatával. Csak a magasabb alapfrekvenciák és női hangosztályok énekeseire, tehát a szoprán és a mezzoszoprán énekesekre koncentrálva összefoglalásunkat a vizsgálat eredményei egyértelműen arra mutattak, hogy a gége függőleges helyzete az alapfrekvencia függvényében nem tekinthető állandónak, és nem igazolódott az a várakozás sem, mely szerint mind a magasabb, mind az alacsonyabb alapfrekvenciákat egyaránt alacsony gégehelyzettel képeznék az énekesek. Ugyanakkor az eredmények alapján a nagymértékű variabilitáson túl nem volt megállapítható semmilyen általános érvényű tendencia az alapfrekvencia és a VLP kapcsolatára. Ráadásul a variabilitás nem is csak az énekesek között mutatkozott (nem arról van szó tehát, hogy minden énekes sajátos, de minden kondícióban következetes tendenciát alkalmazott volna), hanem az egyes énekesek produkcióján belül is: a szerzők ugyanis egymástól gyökeresen eltérő összefüggéseket (vagy az összefüggések teljes hiányát) tapasztalták az  $f_0$  és a VLP között a kétoktávnyi otkávmenet, az öthangnyi skálázás, illetve az oktávugrás énekelési feladatokban. Konklúziójukban a szerzők mindösszesen annyit állítanak, hogy az igazán magas alapfrekvenciáknál minden énekesnél látható volt a gége megemelkedése, valamint kifejtik, hogy eredményeik alapján bizonyítottnak vélik azt az állítást, mely szerint az alacsony

gégehelyzet egyáltalán nem szükségszerű velejárója a professzionális és jó technikával képzett énekhangnak.

Ráadásként a szerzők a hangerősség, illetve hangintenzitás és a gégehelyzet összefüggéseinek vizsgálatára  *messa di voce*  énekfeladatokat is rögzítettek (ekkor az énekes egy adott alaphangfrekvencián fokozatos hangosítást, *crescendót*, majd halkítást, azaz *diminuendót* hajtott végre), valamint a tüdő vitálkapacitásának maximális kimerítését kérték az egyes skáláknál, hogy megfigyeléseket tehessenek a tüdőtérfogat és a gégehelyzet összefüggéseire is. Az első kérdésben a szerzők azt találták, hogy az intenzitás csak néhány énekes esetében volt hatással a gége függleges helyzetére, mégpedig úgy, hogy a halkítás során a gége süllyesztését, a hangosítás során pedig a gége megemelését lehetett tapasztalni. A második kérdésben talált összefüggéseket, melyek szerint nagyobb tüdőtérfogat esetén a VLP alacsonyabb, valamint a VLP egyre nő a tüdőtérfogat csökkenésével (tehát a kilégzéssel) a szerzők a *tracheal pull* ('légcsövi húzás') néven emlegetett jelenség következményeinek vélik. Eszerint mivel belégzéskor a rekeszizom megfeszül, a tüdő húzó hatást fejt ki a felfüggesztésére, mégpedig úgy, hogy lefelé húzza a gégét, különösen a gégeporcokból a hozzá közvetlenebbül kapcsolódó gyűrűporcot, így egyúttal távolítja azt a pajzsporcától. Mivel ez a húzó hatás az alaphangfrekvencia emeléséhez (illetve a magas alaphangfrekvencia megtartásához) szükséges mozgásoknak, azaz a pajzsporc előrebillentésének (és a gyűrűporchoz közelítésének) ellene hat, megnehezíti azt, aminek eredményeképpen a cricothyroideus izom a belégzést követően jobban megfeszül (hogy az alaphangfrekvencia-csökkenést elkerülje). Emellett ráadásul az is igazolt, hogy a húzó hatás miatt (tehát nagyobb tüdőtérfogat esetén) a hangszalagok jobban eltávolodnak egymástól, tehát a tüdőtérfogat egyértelműen hat a fonációra (vö. Sundberg et al. 1988).

Hurme és Sonninen (1995) röntgenképek segítségével vizsgálta mind a gége vertikális (VLP), mind pedig a horizontális (szagittális) elmozdulását (sagittal larynx position, SLP) az alaphangfrekvencia függvényében tizenkét énekesnél, akik közül kilenc tartozott a szoprán vagy mezzoszoprán hangosztályba. Az énekesek a felvételek során az /a/ hangzót ejtették emelkedő skálák hangjain. A korábbiakhoz hasonlóan ez a vizsgálat is igen nagy variabilitást mutatott ki az egyes énekesek között. Míg két szoprán esetében a mozgások összessége egy, az alaphangfrekvencia emelésével hátra (posterior irányba) és felfelé (superior irányba) irányuló elmozdulást



mutatott, négy esetben inkább előre (anterior irányba) és felfelé (superior irányba) mozdult az énekesek gégéje. Egy esetben a mozgások összessége ráadásul inkább hátra (posterior irányba) és lefelé (inferior irányba) mutatott, míg az utolsó esetben egy leginkább rendszertelennek mondható, komplex cikkcakk mozgást tapasztaltak a kutatók. A vizsgálat eredményei szerint tehát – a korábbiaknak megfelelően – a konstans alacsony gégehelyzetre vonatkozó hagyományos feltételezés nem nyert megerősítést, valamint – ismét csak a korábbi eredményekkel egybeeső módon – a szerzők az f0 és VLP, illetve az SLP összefüggéseinek számos lehetséges megvalósulását demonstrálták. Ezen túlmenően a vizsgálat arra is rávilágított, hogy a függőleges gégehelyzet önmagában nem elegendő a „gégehelyzet” leírásához. A lényeg azonban, mint írják, feltehetőleg nem a gégemozgások konkrét útvonala (és összefüggése az f0-lal vagy az intenzitással), hanem az egymásra ható, az izmok által kifejtett erők kiegyensúlyozása.

Amint arra már helyenként utaltunk, a függőleges gégehelyzet kérdése több okból lehet meghatározó, ezért is láthatjuk azt, hogy ilyen fontos és sokat vizsgált kérdés. Az egyik ok, hogy – amint arra már utaltunk –, úgy tűnik, a gégehelyzet és a fonáció között fontos összefüggések húzódnak. Míg Zenker és Zenker (1960), valamint Pabst és Sundberg (1992) a le-süllyesztett gégehelyzettel a hangszalagok nagyobb fokú távolodását (abdukcióját) mutatta ki, Sundberg és Askenfeld (1981) a magas gégehelyzetre nagyobb mértékű közelítésről (addukcióról) számolt be, amely a feszített, préselt fonáció miatt valamivel „veszélyesebb” hangszalagrezgési mód, legalábbis az éneklésben. Így tehát elmondható, hogy a függőleges gégehelyzet mind a fonáció módjára, mind a létrejövő hangszínezetre nagy hatással van. A másik (legfőbb) ok pedig az, hogy a gége emelése rövidíti, süllyesztése pedig hosszítja a toldalékcsovet a garatüreg méretének befolyásolásán keresztül, így a gége föl-le mozgatása hatással van a formánsfrekvenciák értékére is: a gége süllyesztése csökkenti, emelése pedig növeli a toldalékcso sajátrezonanciáinak értékét. A kísérleti eredmények szerint a gége emelése növeli a zárt vagy első nyíltsági fokú palatális hangzók második formánsának értékét, növeli mind az F1, mind az F2 értékét a nyílt vagy harmadik és negyedik nyíltsági fokú hangok esetében, valamint több hangzó esetében a harmadik és negyedik formáns értékének egyidejű növekedését is okozza (Sundberg–Nordström 1976). Ezzel összefüggésben az is valószínűsíthető, hogy a

vertikális gégehelyzetnek a formánsfrekvenciákat befolyásoló hatásával függ össze a férfi hangkategóriák esetében sokat emlegetett énekesformáns megjelenése is.

Az énekesformáns a nevével ellentétben valójában nem egy önálló rezonanciafrekvencia, hanem egy formánsklaszter (egy nagy sávszélességű, nagy intenzitású frekvenciasáv), mely az F3, F4 és F5 formánsokból jön létre. A klaszter létrejöttének oka, hogy a mindennapi beszéd során egyébként nem változtatott (mert a szájüreg artikulátorokkal egyébként nem változtatható) frekvenciaértékű F3, F4 és F5 formánsok frekvenciájukban valamilyen módon közelednek egymáshoz, így (pusztán a csatolt rezgések fizikai törvényszerűségeinek következtében) megemelik egymás amplitúdóját, tehát e formánshelyek frekvenciasávja felerősödik (többlet energiát kap a hang spektrumában) megközelítőleg a 2-2,5 és 3-3,5 kHz-es tartomány között. Ennek eredményeképpen az énekesformánssal bíró énekhang felerősödik abban a frekvenciatarományban, melyben a kísérő zenekar már nem képes maszkolni a hangot (hiszen ezen a frekvenciasávon a hangszerek hangja már nem tartalmaz intenzív összetevőket), így az énekes hangja a zenekari kíséret mellett is jól hallható lesz (Sundberg 1974, 2001). Amint azt a bevezetőben már tárgyaltuk, a szájüregi artikulátorok alapvetően csak az F1 és F2 formánsok frekvenciaértékét módosítják, az F3, F4 és F5 formánsok értékének nagyobb mértékű befolyásolására nem alkalmasak. A függőleges gégehelyzetnek a rezonanciafrekvenciákat befolyásoló hatása azonban Sundberg és Nordström (1976) vizsgálata óta részleteiben is ismert. A vizsgálatok eredményeire alapozva tehát máát régóta élő feltételezés, hogy az énekesformáns kialakítását a függőleges gégehelyzet változtatásával érik el az énekesek, a későbbi kutatás tanulságai szerint egészen pontosan olyan módon, hogy a VLP változtatásával a garatüreg és gégecső arányát megközelítőleg 6 : 1 arányra állítják (Sundberg 1974, Titze 1998, 2001, Sundberg 2001).

Az énekesformáns meglétét azonban csak férfi hangosztályok énekeseinek produkciójában vélik igazoltnak. Sundberg (2001) szerint a női hangosztályok, elsősorban a szopránok hangja e kérdés tekintetében egyáltalán nem mutat egységes képet. Összefoglalóan azt írja, hogy bár némely vizsgálat sikerrel mutatott ki valamifajta intenzitáserősödést a magasabb frekvencia-komponensekben a szopránok esetében is, ennek az erősödésének a mértéke azonban nem volt a férfi hangosztályokban tapasztaltakhoz mérhető. Mindemellett feltételezhető, hogy ez az in-

tenzitáserősödés hozzájárul a női énekhang hallhatóságának növeléséhez nagy terekben és nagyzenekari kíséret mellett (vö. Sundberg 1977, Sundberg 2001, Thorpe et al. 2001, Weiss et al. 2001). Mindehhez még egy további megjegyzés is fűzhető. Az énekesformáns vizsgálata a szoprán hangokban eleve sokkal nehezebb feladat, mint a férfi hangokban, ráadásul egyszerre két okból: egyrészt azért, mert az alul-mintavételezés miatt a felsőbb formánsok detekciója is akadályozott, másrészt pedig azért, mert az  $F1 : f_0$  hangolás következtében a magas női énekhangok spektrumának energiája már eleve nagyobb, tehát az  $F1 : f_0$  hangolása és az énekesformáns kialakítása a magasabb formánsokból gyakorlatilag disszociálhatatlan. Ennek értelmében azonban feltehetőleg nincs is értelme (vagy relevanciája) az énekesformáns kérdésének a női hangosztályok esetében, hiszen valószínűsíthető, hogy nem ez az eszköz segíti hozzá a női szólistákat a zenekaron is átható, hangos énekprodukciónak, hanem sokkal inkább az  $F1 : f_0$  hangolás. Mindezen okokból tehát a jelen értekezésben nem foglalkozunk az énekesformáns kérdésével a szopránok esetében.

#### **1.2.1.2. A gége belső mozgásai: a nyelvtőcsont szerepe az éneklésben**

Hurme és Sonninen (1995) a gége helyzetével párhuzamosan vizsgálták a gégeporcok egymáshoz viszonyított helyzetét, valamint a nyelvtőcsont helyzetét is, tehát a gége egyes belső mozgásait. Míg a nyelvtőcsont szerepe a beszédben és az ember élettani működései szempontjából a mai napig némiképp rejtélyes (vö. Seikel et al. 2010), addig Hurme és Sonninen (1995) az éneklésben kapott adatai – kiegészítve Honda a témában folytatott vizsgálataival vö. pl. Honda (1983) – e csontocska egy (vagy több) lehetséges funkciójára is utalnak.

Hurme és Sonninen (1995) azt találta, hogy a magasabb alapfrekvenciákat nagy arányban magasabb gégehelyezettel produkáló énekeseknél a nyelvtőcsont a magas alapfrekvenciákon a beszédben nem tapasztalható mértékben a pajzsporc elé billent, ráadásul olyannyira, hogy annak elülső, legalacsonyabban elhelyezkedő pontja akár a pajzsporc csúcsát (azaz az ádámcsutkát) is elérhette. Bár a nyelvtőcsont mozgását az  $f_0$ -lal szoros összefüggésben figyelték meg, a szerzők úgy vélik, hogy ezeket a mozgásokat elsősorban nem az  $f_0$ -kontroll, hanem a VLP változása befolyásolja. Következtetésük szerint ugyanis a megemelkedő gégehelyzet esetén a

levegő útja akadályoztatottabb lehet, amit a nyelvtőcsont előrebillentése kompenzálni tud, mert hatására a gégecső tágulni tud.

Egy másik elképzelés szerint azonban a nyelvtőcsont az  $f_0$  befolyásolásával van szoros összefüggésben. Honda (1981, é.n.) vizsgálatai szerint ugyanis az  $f_0$  extrém emelésekor (amire például énekléskor van szükség) a nyelvtőcsonthoz kapcsolódó izmok (többek közt a geniohyoideus vagy a genioglossus) is megfeszülnek, és előre húzzák a nyelvtőcsontot, azért, hogy az egyébként normál körülmények között (de kisebb hatásfokkal) az alapfrekvenciát emelő cricothyriodeus izom működését, a pajzsporc előrebillentését fokozzák. Ha ugyanis a nyelvtőcsont előrebillen, és lefelé húzódik, a nyelvtőcsont és a pajzsporc között lévő rugalmas szövetek (a thyrohyoideus izom, a középső és szélső thyrohyoideus szalagok, a thyrohyoideus membrán stb., vö. 2. ábra) segítségével a nyelvtőcsont magával húzza a pajzsporcot is. Így a pajzsporc a normál rotációs fokánál nagyobb szögben képes elmozdulni, így jobban megfeszítheti a hangszalagokat is, tehát a nyelvtőcsont előrebillentésével magasabb alapfrekvenciák érhetők el (Honda 1983, é.n.). Mindez azonban egy komplex rendszer része, a nyelvtőcsont pontos kontrolljához ugyanis nem csak az azt előrébb mozdító geniohyoideus vagy genioglossus izmok megfeszítése szükséges, hanem az azok antagonista (ellentartó) izmait képező, a nyelvtőcsontot hátra, a garat felé húzó középső garatösszeszorító izom (constrictor pharyngis medius) és a nyelvtőcsontot lefelé húzó sternohyoideus pontos működtetése (Honda é.n.).

A nyelvtőcsont helyzetét és funkcióját vizsgáló kutatások tehát részben választ adnak egy, az énekprodukciónak kapcsolatban talán alapvetőnek tűnő kérdésre, mégpedig arra, hogy hogyan képesek a szoprán énekesek az énekmodszertanilag nem képzett beszélőkhöz képest jóval magasabb és jóval kontrolláltabb alapfrekvencia-emelésre. Az énekesek ugyanis feltehetően nemcsak a normál beszéd módban (és egyébként minden beszélő által) használt cricothyroideus izom működtetésével emelik és kontrollálják hangjuk alapfrekvenciáját, hanem a nyelvtőcsont összeköttetéseit kihasználva a fent felsorolt komplex rendszer egészének begyakorlott manipulálásával. Ráadásul a fenti eredmények mindeközben arra is rávilágítanak, hogy hiába elegendő bizonyos kérdések megértéséhez és leképezéséhez a beszédképzés lineáris, a forrást és a szűrőt, tehát a gége és a toldalékcsovet élesen elválasztó modellje (melyben

a gége működésére elvben nem hat a toldalékcső tevékenysége), a valóságban a beszédképző rendszer a forrás-szűrő modell által sugalltnál jóval összetettebb módon működik, a rendszer részei egymást közvetlenül vagy közvetetten befolyásolva vesznek részt az emberi hangadás folyamatában.

### **1.2.2. Az énekelt magánhangzók akusztikai vizsgálata, rezonanciák a szoprán éneklésben**

Az énekelt magánhangzók akusztikai megvalósulásának esetében, amint már említettük, a fő kérdés, hogy mi történik akkor, ha az  $f_0$  magasabbra emelkedik, mint a beszédben jellemző  $F1$  értéke ( $F1_{\text{beszéd}}$ ). Mivel a szopránok és a szoprán énekesekre írt művek hangetrjedelme átlagosan a  $c'$  (262 Hz) zenei hangtól a  $c'''$  (1047 Hz) zenei hangig terjed, énekprodukciónk során a szopránok igen gyakran énekelnek ehhez fogható magas alapfrekvenciákon<sup>11</sup>. Tudjuk, hogy ha az  $F1$  az  $f_0$  értéke alá kerülne, tehát az énekes nem változtatna a magánhangzónak a beszédben megszokott ejtésén a megemelt alapfrekvencián, az énekesnek három nagyon fontos, az énekelt hangzók nyelvi és esztétikai minőségét is befolyásoló következménnyel kellene számolnia. Az egyik következmény egy a magánhangzó-minőség kialakítása és észlelete szempontjából igen fontosnak tartott (l. pl. Gósy 2004) akusztikai kulcs, az  $F1$  elvesztése, a másik az akusztikai energia, a hangnyomásszint, tehát a hangosságérzet csökkenése, a harmadik pedig a hangszínből hirtelen beálló változás lenne (vö. Sundberg 1979, 1987). Ennek megfelelően a szopránok ejtéséről azt feltételezik, hogy (magas alaphangokon) az alapfrekvencia függvényében változik, mégpedig úgy, hogy ha az  $f_0$  értéke magasabb, mint az  $F1_{\text{beszéd}}$ , a szoprán az  $F1$  értékét felhangolja a megemelt  $f_0$  értékére vagy valamivel afölé (l. pl. Sundberg 1987).

Sundberg (1979, 1987) szerint, aki a szoprán énekesek alapfrekvenciafüggő produkciós tendenciáit elsőként tanulmányozta mélyrehatóan, az  $F1 : f_0$  hangolásról szóló feltételezés empirikus bizonyítékok nélkül is igen stabil lábakon állna, hiszen tudható, hogy már pusztán

---

<sup>11</sup> A zárt képzésű magyar magánhangzók első formánsa a toldalékcső hosszától, a beszédstílustól és beszédtempótól is függően nagy variabilitást mutathat, de átlagosan 200 és 500 Hz közötti értéken realizálódik (l. pl. Gráczy–Horváth 2010 és a jelen dolgozat 2.1. és 2.5. alfejezete).

az F1-nek az f0-ra hangolásával, azaz a megszólaló hang alapfrekvenciájának felerősítésével 20 vagy akár 30 dB hangnyomásszint-növekedés (azaz hangerő-növekedés) is elérhető. A formánshangolás, azaz a toldalékcso rezonanciatevékenységének jobb kihasználása ráadásul úgy idézi elő a hangerőnövekedést, hogy ahhoz semmilyen további vokális erőfeszítés (pl. a szubglottális nyomás növelése, vö. kiabálás) nem szükséges (l. még Titze 1994). Sundberg szerint, mivel az operaénekesek hagyományosan minden esetben erősítés nélkül énekelnek, könnyen belátható, hogy nagyon sokat profitálhatnak abból, ha énektechnikájukat ennek az akusztikai törvényszerűségnek a kihasználásával, azaz az üregi rezonanciák jobb kihasználásával fejlesztik. A felvetés alátámasztása empirikus bizonyítékokkal azonban nehéz, hiszen a magas alapfrekvenciákon ejtett magánhangzók spektrálisan alul-mintavételezettek, azokban a formánsfrekvenciák középtérének meghatározása a hagyományos, Fourier-transzformáción alapuló elemzéssel nem megbízható (vö. de Cheveigné–Kawahara 1999, valamint a jelen értekezés *1.1. alfejezete*).

Éppen ezért az egyes nyelvek magánhangzóinak minőségét (azaz akusztikai értelemben az első két formánsukat) a magas alapfrekvenciájú éneklés során érintő változásokról szóló információink forrása kettős. A vizsgálatok egyik (és talán valamivel nagyobb) része közvetett, mert ezek elsősorban az énekelt magánhangzók artikulációját vizsgálták. E vizsgálatok némelyike ugyanakkor a puszta artikulációs adatok mellett az énekelt magánhangzók formánsfrekvenciáit is igyekezett megállapítani, mégpedig az artikulációs adatokból kalkulált fizikai és matematikai modellek segítségével. A vizsgálatok másik csoportja pedig, noha némileg szintén közvetett módon, de kizárólag akusztikai elemzések révén igyekezett meghatározni az énekelt magánhangzók formánsértékeit. Ez utóbbi vizsgálatok a magas alapfrekvenciájú hangforrásokra jellemző alul-mintavételezettség miatt a hagyományos, Fourier-transzformáción alapuló akusztikai elemzésektől eltérő, újszerű kísérleti méréseket jelentettek. A következőkben ezeket a magas alapfrekvencián énekelt magánhangzókat a direkt és indirekt módszerekkel vizsgáló kísérleteket és a vizsgálatok eredményeit tekintjük át. Azokban az esetekben, ahol egy vizsgálaton belül a kutatók artikulációs (és azokból derivált akusztikai) és ezen felül valamely akusztikai adatokra támaszkodva is megfigyeléseket tettek, a kettőt szétválasztva közöljük, hogy az

egyes eredmények miként (direkt vagy indirekt bizonyítékokként) értelmezendők a toldalécső rezonanciáira nézvést.

A szoprán énekesek a formánshangolást előidéző artikulációs stratégiáival kapcsolatban Sundberg (1975), Johannson és munkatársai (1982), Sundberg és Skoog (1995), valamint Bresch és Narayanan (2010) folytattak vizsgálatokat (rendre) egy különleges, egyéni készítésű, a metszőfogak és a két állkapocs távolságát mérő mérőeszköz (vö. Cooker 1973), röntgenográfia, egy magnetométer alapú mérőeszköz (vö. Branderud et al. 1994), valamint mágneses képalkotó eljárás (MRI) segítségével.

Sundberg (1975) a svéd /u o a e i y/ magánhangzókat vizsgálta 262 Hz, 394 Hz, 523 Hz és 698 Hz alapfrekvencián (a *c'*, *g'*, *c''*, *g''* zenei hangokon) énekelve, egy szoprán ejtésében. Eredményei szerint, míg az /a/ ejtésékor az *f*<sub>0</sub> emelését közel állandó mértékű állapcsonyítás kísérte, addig a többi vizsgált hangzó esetében az *f*<sub>0</sub> emelésével párhuzamosan a nyitásszög fokozódása volt tapasztalható. Az adatokból Sundberg arra következtetett, hogy a kísérletben részt vevő énekes az énekelt magánhangzóinak első formánsát az alapfrekvencia és a magánhangzó-minőség függvényében egyre feljebb és feljebb hangolta az *f*<sub>0</sub> emelésével, ha az *f*<sub>0</sub> elérte vagy megközelítette az *F*<sub>1</sub> értékét. Az említett magyarázatra okot az a mára már közhelyként összefüggés szolgáltatott, mely szerint az állkapocsnyitás szögének növelésével az első formáns frekvenciaértéke is növekszik (vö. pl. Lindblom–Sundberg 1971). Sundberg szerint az /a/-ra talált kivételes tendencia (az /a/ ejtésékor az *F*<sub>1</sub> az *f*<sub>0</sub>-tól látszólag független viselkedése) is az *F*<sub>1</sub> : *f*<sub>0</sub> hangolás feltételezésének helytállóságát bizonyítja, hiszen az /a/ nyílt (negyedik nyíltsági fokú, legalsó nyelvvállású) hangzó, mely a beszédben is magas *F*<sub>1</sub>-értékkel realizálódik. Így tehát a kísérletben vizsgált alapfrekvenciákon az *F*<sub>1</sub> : *f*<sub>0</sub> hangolási stratégia feltételezése mellett is az várható, hogy az /a/ frekvenciaszerkezete változatlan marad az *f*<sub>0</sub> emelésével.

Johannson és munkatársai (1982) a magánhangzótér artikulációs és akusztikai értelemben is szélső értékeit képező /a i u/ megvalósítását vizsgálták egy szoprán és egy alt énekesnél, a szoprán esetében 230 Hz, 470 Hz és 950 Hz alapfrekvenciákon, míg az alt esetében 150 Hz, 300 Hz és 600 Hz alapfrekvenciákon. A vizsgálat során a kutatók az ajak- és állkapocsnyitás szögét, valamint a gége függőleges helyzetét mérték laterális röntgenfelvételek segítségével,

továbbá a toldalékcső középvonalaának meghatározása után kiszámolták a toldalékcső adott pontjain jellemző keresztmetszeteket (azaz a toldalékcső keresztmetszetfüggvényét) és a keresztmetszeti területeket (azaz a toldalékcső keresztmetszetiterület-függvényét) is. A toldalékcső egyes szakaszaira jellemző keresztmetszeti területeket az egyes szakaszokra meghatározott keresztmetszetek konverziójával kapták (melyhez a hiányzó adatokat, azaz a röntgenfelvétel síkjára merőleges távolságokat a garatüregre és a szájüregre jellemző, korábban meghatározott állandók pótolták). A keresztmetszetiterület-függvény alapján a kutatók plexilapokból fizikai csőmodellt készítettek, melyet szinuszhullámokkal gerjesztve meg tudták határozni a cső sajátfrekvenciáit, azaz a csőmodell alapját képező toldalékcső-konfigurációra jellemző formánsértékeket is.

Johannson és munkatársainak (1982) az F1 frekvenciaértékének manipulációja szempontjából fontos artikulációs eredményei szerint a szoprán résztvevő következetesen növelte az állkapocsnyitás szögét az  $f_0$  emelésével, valamint mind a nyelvkontúr, mind az ajkak távolsága, mind pedig az ajkak „behúzása” tekintetében uniformizálta az /a/, /i/ és /u/ hangzóit a legmagasabb, 950 Hz alapfrekvencián. Az akusztikai eredmények azonban azt mutatták, hogy ugyananezen énekes nem az  $f_0$  értékére, hanem valamivel az  $f_0$  értéke alá hangolta az F1-rezonanciát akkor, amikor az  $f_0$  egyébként magasabb értéket vett volna fel, mint az F1 beszédben jellemző frekvenciaértéke. Ez utóbbi eredményt a szerzők a modellből fakadó hibaként értelmezték, és értelmezésük szerint az artikulációs adatok egyértelműen bizonyították az F1 emelésére, valamint az F1 :  $f_0$  hangolásra irányuló produkciós tendenciákat. Mindezen felül a szerzők arról is beszámoltak, hogy a szoprán énekes (szintén az  $f_0$  emelésével) csökkentette az elől képzett, és növelte a hátul képzett hangzók második formánsának értékét is, így uniformizálva a három beszédhangot a kísérletben vizsgált legmagasabb alapfrekvencián, azaz 950 Hz-en.

Sundberg és Skoog (1995) a svéd /a a o u i e/ magánhangzókat vizsgálta egy szoprán és három mezzoszoprán (továbbá egy mezzoszoprán-alt, két tenor, két bariton és egy basszus énekes) ejtésében, két oktáv hangterjedelemben, mely a szoprán énekesnő esetében a 220 és 880 Hz (azaz az *a* és *a''* zenei hangok) közötti, a mezzoszopránok esetében pedig a 196 Hz és 698 Hz (azaz a *g* és *f''*) közötti alapfrekvencia-tartományt jelentette. A két kutató vizsgálatában a



szopránokat és mezzoszopránokat érintő eredmények tekintetében a korábbiakkal csak részben egyező eredményekre jutott, ugyanis az egyes magánhangzókra (valamint az egyes hangosztályok képviselőinek stratégiáira) eltérő tendenciákat mutattak ki a következők szerint. Úgy találták, hogy a magasabb beszédbeli F1-gyel realizálódó /a α/ hangoknál egyértelműen tapasztalható volt az állkapocsnyitás szögének növelése akkor, amikor az f0 megközelítette az F1 értékét. A beszédben valamivel alacsonyabb F1-gyel realizálódó /o e/ esetében azonban ez a tendencia csak magasabb alapfrekvencián, mégpedig körülbelül 5 félhanggal az után kezdődött, hogy az f0 elérte volna az F1 értékét az f0 emelése során. A beszédben legalacsonyabb F1-gyel realizálódó /u i/ esetében találták a szerzők a legkomplexebb és legdivergánsabb eredményeket. Az /u/ és /i/ hangok esetében ugyanis azt tapasztalták, hogy minden énekes, köztük a szoprán és mezzoszoprán énekesek is csak az általuk a vizsgálatban lefedett frekvenciatartomány legfelső részében mutattak az f0-tól függő, az állkapocsnyitásszög növelésére irányuló produkciós tendenciát. A mezzoszopránok esetében az állkapocsnyitásszög növelését célzó tendencia az F1 beszédbeli értéke fölött 7 félhanggal, míg a szoprán énekes esetében az F1 frekvenciaértékéhez közel (körülbelül az F1 frekvenciaértékén) kezdődött az f0 emelése során. (Érdekességgéppen megemlítjük, hogy ugyanezen vizsgálatban az /u i/ hangok esetében a tenorokra a szopránokhoz hasonló tendenciát találtak, viszont a bariton és basszus hangosztályú énekeseknél azt látták, hogy az állkapocsnyitásszög növelése már a beszédben jellemző F1 értéke alatt elkezdődött.)

Sundberg és Skoog (1995) végezetül két következtetést fogalmazott meg. Az egyik szerint: mivel az állkapocsnyitás nemcsak az F1 értékének módosítására van hatással, hanem a fonációra is, valójában nem meglepő az, hogy nem minden magánhangzó és nem minden énekes hangosztály esetében figyelhető meg az állkapocsnyitás az f0 értékétől függő növelésének uniform tendenciája. A tanulmány másik konklúziója pedig valójában egy, a jövőben tesztelendő feltételezés megfogalmazása: a szerzők szerint ugyanis a beszédben alacsonyabb F1-gyel realizálódó (zárt, első nyíltsági fokú vagy felső nyelvvállású) magánhangzók esetében valószínűsíthető, hogy azok első formánsának értékét nem az állkapocs, hanem más artikulátorok (mégpedig a nyelv) manipulációja révén módosítják az énekesek, elsősorban kényelmi szempontokból. Magyarázatuk szerint ugyanis, míg a nyíltabb ejtésű magánhangzók

esetében a nyelv által okozott szűkület csökkentése (ami az állkapocsnyitásszög-módosításnál ezen a nyelvállásfokon ökonomikusabb manipuláció) a kívánatossal éppen ellenkező hatású lenne (hiszen csökkentené az F1 értékét), a zártabb /u i/ hangoknál (melyek a nyelvhelyzet módosítására a nyitottabb hangzóknál jóval érzékenyebbek) ez a módosítás a kívánt eredményt, azaz az F1 emelését hozza (erre a felvetésre hivatkozási alapot Lindblom–Sundberg 1971 eredményei adnak). Így bár Sundberg és Skoog (1995) artikulációs eredményei közvetlenül nem igazolják a szopránoknak az F1 hangolására tett erőfeszítéseit az /u/ és /i/ hangok esetében, valamint kétségeket hagynak a hangolás megjelenését illetően az /o/-ra és /e/-re nézvést is, a talált tendenciákat a kutatók az F1 az f0 emelését kísérő manipulálása közvetett bizonyítékként értelmezik, ismét csak az állkapocsnyitás mértékének és az első formáns frekvenciaértékének összefüggéseire alapozva.

Bresch és Narayanan (2010) öt szoprán énekes énekprodukciónak vizsgálta két oktáv hangterjedelemben (233 Hz és 932 Hz, azaz az *a*# és az *a*#'' zenei hangok között), az /a e i o u/ hangokon skálázva valós idejű mágneses rezonancián alapuló képalkotó eljárás (MRI) segítségével, vibráló nélküli énekmódban. Bresch és Narayanan az MRI segítségével készített szagittális felvételekből az énekelt magánhangzók artikulációját, valamint akusztikai megvalósulását is vizsgálta. A szerzők az akusztikai mérésekhez az oldalnézeti képek segítségével megalkották a toldalékcső csőmodelljét, majd matematikai úton kiszámították a csőhöz tartozó formánsfrekvenciákat. A modell létrehozásához előbb kiszámították az artikulációs csatorna egyes szakaszaira jellemző oldalnézeti keresztmetszeteket (ezeket az egyes szakaszoknak a gégtől való távolságára vetítve megkapták a toldalékcső keresztmetszetfüggvényét, aperture function), majd ezeket a keresztmetszetfüggvényeket keresztmetszeterület-függvényekké (cross-sectional area function) konvertálták, végül pedig a csőmodellből (bár nem fizikai, hanem matematikai úton, de) Johansson és munkatársainak (1982) vizsgálatához hasonló módon meghatározták az üregi rezonanciákat.

Artikulációs adataik alapján Bresch és Narayanan (2010) megállapították, hogy az f0 emelésével minden énekes egyre nyitottabb és egymástól egyre kevésbé eltérő artikulációs beállítással ejtette az öt különböző minőségű magánhangzót (a nyitódást elsősorban a szájüreg elülső részére vonatkoztatva). A modell alapján számított akusztikai adataikat (az F1 értékét) az f0

értékével összevetve pedig a szerzők azt találták, hogy csak az /i/ és /u/ F1-e esetében lehetett látni minden énekes esetében szignifikáns korrelációt – az /e/ és /o/ esetében már csak két énekesre kaptak szignifikáns összefüggéseket, míg az /a/ esetében egyetlen adatközlő F1 és f0 értékeire sem.

Bresch és Narayanan eredményeiből arra következtetett, hogy csak az /i/ és /u/ esetében mondható ki egyértelműen az F1 az alapfrekvenciától függő általános (tehát nem énektechnika- vagy énekesfüggő) hangolási stratégiája. A kutatás eredményeinek értelmezéséhez azonban még két tényezőt fontos látnunk. Az egyik – és ezt részben a szerzők is hangsúlyozzák –, hogy az /a/ esetében a korreláció, azaz a hangolási tendencia hiánya várt eredmény. Amint a korábbi vizsgálatok is egybehangzóan tanúsították, az /a/ és az /a/ ejtése stabilabb, az alapfrekvenciától nagyobb mértékben független, mert ezek nyitottabb ejtésű, magas első formánssal rendelkező hangzók, tehát az F1 formánst csak magas alapfrekvenciák éneklésekor ér el az f0 az f0 emelése során. A másik tényező pedig az, hogy nyíltsági fokukból kifolyólag valójában az /e/ és /o/ esetében is csak a magasabb, 4-500 Hz fölötti alapfrekvenciákon várnánk az F1-nek a beszédbeli átlagos frekvenciájánál magasabb frekvenciájú megjelenését, hiszen az /e/ és /o/ hangok F1-értéke is átlagosan 4-500 Hz körül jelenik meg. Ez Bresch és Narayanan vizsgálatára nézve azt jelenti, hogy az ott vizsgált alapfrekvenciák alsó tartományában, azaz kvázi az első oktávon és még valamivel a fölött sem várnánk hangolást az /e/ és /o/ esetében sem, így ezen a tartományon korrelációt sem prediktálhatunk az F1 és f0 értékek között. Következésképpen feltételezhetjük, hogy a 16, a korrelációs számításnál vizsgált adatpont (azaz a 16 alapfrekvenciaérték és az ahhoz tartozó F1-érték) második felének összefüggéseit, azaz a második, magasabb oktávban talált hangolási tendenciákat elfedhették a hangolás szempontjából nem kiritikus tartományon (azaz az első oktávon) talált tendenciák. Ilyen módon tehát Bresch és Narayanan (2010) eredményei valójában nem az F1 : f0 formánshangolási stratégia hiányát bizonyították az egyes hangzók esetében, hanem csak azt az elképzelést erősítették meg, mely szerint az F1 : f0 hangolás megjelenése az F1 frekvenciaértékétől, tehát a hangok nyíltsági fokától függ.

A szerzők a következtetések között azt is megjegyezték, hogy a vizsgálatban kapott (a modellből számított) formánsfrekvenciák értékének pontosságával kapcsolatban aggályaik merül-

tek fel – úgy vélték ugyanis, hogy az általuk mért formánsfrekvenciák a feltételezett valós frekvenciáknál (a Johansson és munkatársai 1985 végezte vizsgálathoz hasonlóan) alacsonyabbak voltak.

A direkt akusztikai adatokkal dolgozó vizsgálatok egyik és talán legfontosabb eltérése a hagyományos akusztikai elemzésekhez képest egy a beszélőtől független, a természetes fonációt helyettesítő külső gerjesztés, külső forrásjel alkalmazása. Az itt bemutatásra kerülő vizsgálatok során ugyanis a kutatók egy nagy felhangsűrűségű, alacsony alapfrekvenciájú külső gerjesztésű jelet juttattak az artikulációs csatornába (technikától függően a nyaknál, a gége fölött vagy a szájon keresztül), majd a toldalékcsőből kijutó jelet a szájnál (mikrofon segítségével) rögzítették. Mivel a külső forrás lehetővé teszi, hogy a toldalékcső átviteli karakterisztikáját a fonációtól független, alacsony alapfrekvenciájú, nagy felhangsűrűségű hang segítségével „újrámintavételezzük”, ezekben a vizsgálatokban a külső gerjesztés használata által kikerülhetővé vált a magas alapfrekvenciájú hangforrás elemzésének problémája. A külső gerjesztésű, de a toldalékcső által megszűrt hang elemzésével ugyanis mérhetővé válnak a magas alapfrekvencián ejtett hangok formánsfrekvenciái is, melyek a természetes hangadásban a magas alapfrekvencia és a felhangok közti nagy távolság miatt nem állapíthatók meg a hagyományos, Fourier-transzformáción alapuló elemzési eljárásokkal.

Sundberg (1975) már említett vizsgálatában nem csak artikulációs, hanem akusztikai méréseket is végzett: a svéd /u o a e i y/ hangzókat elemezte, 262 Hz és 698 Hz alapfrekvenciákon (a zenei *c'* és *g''* hang) között egy szoprán ejtésében, úgy, hogy az egyes hangok ejtésekor az énekes zöngéjét egy elektromos gége gerjesztette forrásjellel helyettesítette. A felvételek a következőképpen zajlottak. Az énekes először saját fonációjának segítségével produkálta az énekelt magánhangzókat, majd bekapcsolt egy, a gégéje felett (a garat magasságában) a nyakához illesztett elektromos gégét, mely így egy mély alaphangú, nagy felhangsűrűségű berregő hangot juttatott a toldalékcső alsó részébe. Ezután egy adott pillanatban az énekes abbahagyta zöngképzést (ekkor már csak az elektromos gége keltette berregő hang jelentette a száját elhagyó kimeneti jel forrását), és bezárta a hangrést (a levegő visszatartásával). Mivel a fonáció abbahagyásakor az énekes igyekezett megmerevíteni az artikulációját (és még pár másodpercig „némán énekelve” tartani azt), az ekkor a száját elhagyó nagy felhangsűrűségű, az elekt-

romos gége keltette akusztikai jelben az artikulációs konfiguráció aktuális sajátfrekvenciái emelkedtek ki – és váltak mérhetővé (a szerző tehát az elektromos gége alapvető működési elvét használta ki).

A vizsgálatában kapott rezonanciaadatok segítségével Sundberg bizonyítékkal szolgált az F1-nek az f0-ra hangolásának a hangzók nyíltsági fokától, illetve első formánsának értékétől függő tendenciájára a kísérletben vizsgált /u o a e i y/ magánhangzók esetében. Ennek megfelelően Sundberg azt is megállapította, hogy az /a/ formánsai a többi magánhangzónál jóval nagyobb függetlenséget mutattak az alapfrekvenciától, Sundberg feltételezése szerint éppen azért, mert az /a/ nyílt ejtésű, és magasabb első formánssal rendelkezik.

Hertegård és Gauffin (1993) akusztikai vizsgálata csak az /a/ hangra koncentrált 250, 390 és 750 Hz alapfrekvencián. Vizsgálatuk során a kutatók a résztvevő szoprán énekes hangját először egy Rothenberg-maszk, majd egy mikrofon segítségével is rögzítették. A Rothenberg-maszk egy hagyományos mikrofonhoz képest nem a hangnyomás-idő függvényt, hanem a beszéd közben a szájat elhagyó levegőáram mennyiségét (illetve a levegő térfogat-sebességét) regisztrálja egy úgynevezett áramlási glottogram függvényben. Az áramlási glottogram hullámformája tehát a levegő áramlását milliméter/másodperc [mm/s] mennyiségként rögzíti az idő függvényében. Az elemzések során Hertegård és Gauffin a két rögzített hullámforma, azaz a hangnyomás-idő függvény és az áramlási glottogram inverz szűrését végezte el. Mivel az inverz szűrés során az elemző arra törekszik, hogy „kioltva” az áramlási függvényben az üregi rezonátortevékenység okozta energiamaximumokat „rekonstruálja” a transzglottális levegőáramra még a glottisznál jellemző áramlási (térfogat-sebességi) függvényt (bővebben lásd még a jelen értekezés 2.5.3. *alfejezetét*), a szűrés során egyúttal megállapítja a toldalékcső szűrőtevékenységét jellemző rezonanciahelyeket is. Ezt használta ki Hertegård és Gauffin (1993) is: az inverz szűréssel a szerzők nem is elsősorban a hangrésnél jellemző áramlási glottogramot, hanem a glottogramot az éneklés során befolyásoló (megszűrő) toldalékcsői sajátrezonanciákat, azaz a formánsfrekvenciákat becsülték meg.

Eredményeik szerint a szoprán énekes az alaphang emelésével növelte az F1 értékét úgy, hogy az megközelítse az f0 frekvenciaértékét, ha az f0 az F1 beszédben mérhető értékénél magasabbra emelkedett (az énekes tehát hangolta az F1 formánst). Ugyanakkor a szerzők azt

is kimutatták, hogy a mikrofonnal rögzített hanganyagon a Rothenberg-maszkos felvételekből detektálhatóknál magasabb F1-értékek voltak mérhetőek. A két jel szűrése eredményeként előállt formáns-frekvenciabeli különbségeket azzal magyarázták, hogy a maszk (zárt formájából adódóan) korlátozta az énekest állkapcsa szabad mozgatásában (az állkapocs szabad nyitásában) – mely feltételezésüket, leírásuk szerint, az énekes beszámolója is megerősítette.

Joliveau és munkatársai (2004) is külső gerjesztésű forrásjelet alkalmaztak akusztikai mérésekben, ám ők ezt a forrásjelet az énekes szájnyílásánál juttatták a toldalékcsőbe, úgy, hogy besugárzás közben az énekes folyamatosan énekelt. A felvételek során a toldalékcsőből visszaverődött jelet egy mikrofon segítségével rögzítették, majd „megtisztították” az énekes fonációja gerjesztette hangtól, és végül akusztikailag elemezték – azaz meghatározták a jelbesugárzás alatt jellemző artikulációs konfiguráció formánsfrekvenciáit. A szerzők ezzel a módszerrel vizsgálták az ausztrál angol /ɑ ɔ u ʊ/ hangokat (melyeket a *hard*, *hoard*, *who'd*, *heard* hangsorokkal elicitáltak). A kilenc szoprán résztvevő halkan (*piano* módban) és vibrató nélkül (*senza vibrato*) énekelt a vizsgált magánhangzókat kitartva (körülbelül 4 s hosszúságban), az egyes énekesek számára kényelmes alapfrekvenciákon (maximálisan körülbelül 1046 Hz-en). A szerzők a vizsgálatban kimutatták az F1-nek<sup>12</sup> az f0-ra hangolását olyan esetekben, amikor az f0 megközelítette az F1 beszédben mért értékét. Azt találták azonban, hogy bár a hangolási tendencia még megközelítőleg 1000 Hz alapfrekvenciáig jellemző volt az ajakréses /ɑ/ és /ɔ/ magánhangzókra, az ajakkerekítéses /ɔ/ és /u/ esetében 1000 Hz alapfrekvencián már nem jelent meg (itt az F1 az f0-nál alacsonyabb értéken volt mérhető). Noha a szerzők értelmezése szerint az pusztán az F1 hangolásából következik (annak elkerülhetetlen velejárója), tehát nem szándékos stratégia, de Joliveau és munkatársai (2004) az előbbieken felül azt is feljegyezték, hogy az F1 hangolásának megjelenésével az F2 értéke is változott. A jelenség magyarázata a szerzők szerint az, hogy az F1 emelésének eléréséhez szükséges artikulációs változtatások szükségszerűen az F2 értékére is hatással vannak.

---

<sup>12</sup> Az idézett tanulmány élesen elkülöníti az üregi rezonanciákat a fonáció mintavételezte kimeneti jel intenzitásmaximumaitól. Ennek értelmében az előbbiekre az R1, R2 stb. notációt, míg utóbbiakra az F1, F2 stb. notációt alkalmazza. Mivel a jelen értekezésben nem követjük ezt a jelölési konvenciót, és formánsnak az üregi rezonanciákat tekintjük (akkor is, ha azok nem esnek egybe a kimeneti jel intenzitásmaximumaival), ezért mind Joliveau és munkatársainak (2004), mind pedig Garnier és munkatársainak (2010) a rezonanciákra vonatkozó eredményeit F1, F2 stb. megjelöléssel hivatkozunk.

Garnier és munkatársai (2010) ugyanezzel a szájnyílásnál a totalécsőbe juttatott külső gerjesztést alkalmazó módszertani megoldással elemezték az /a/ magánhangzót 440 Hz alapfrekvencia fölött 12 énekes (négy profi, négy haladó és négy kezdő) ejtésében. Eredményeik szerint az énekesek az  $f_0$  értékéhez hangolták az /a/ első formánsát 1000-1200 Hz (a  $c'''$  és  $d'''$  zenei hangok) alatt a képzettségi fokuktól nagyjából függetlenül. 1200-1400 Hz fölött (2300 Hz-ig) azonban az  $F_1 : f_0$  hangoláson felül egy további hangolási tendenciát találtak. Ezen a kifejezetten magas alapfrekvencia-tartományon ugyanis öt énekes az  $F_2$  értékét emelte az  $f_0$  értékére (vagy annak közelébe) – míg volt, aki továbbra is az  $F_1 : f_0$  hangolást is alkalmazta. A szerzők az 1200 Hz fölötti extrém alapfrekvenciákra talált eredményeikkel és az  $F_2 : f_0$  hangolási stratégia létezésével tartják magyarázhatónak azt, hogy a szopránok képesek az ezen a frekvencián már igen kis hangerejű zöngével is a kívánalmaknak megfelelően nagy hangerejű hangadásra.

Összegzőképpen tehát az ismertetett vizsgálatok a svéd és az ausztrál angol nyelv bizonyos magánhangzóit érintették – illetve több esetben csak az /a/ hangzót a nyelv megjelölése nélkül. A vizsgálatok fő kérdése a formánshangolási stratégiák megjelenése volt (különösen az  $F_1$ -et érintő hangolási stratégiák tekintetében). A vizsgálatokból származó (direkt és indirekt akusztikai, illetve artikulációs eredmények) gyakorlatilag egybehangzóan kimutatták az  $F_1$ -nek az alapfrekvenciára hangolását akkor, amikor az  $f_0$  emelésével annak értéke megközelítette az  $F_1$ -nek a beszédben az adott magánhangzóra jellemző frekvenciaértékét. Két tanulmányban kivételként említették a szerzők az /i u/ magánhangzókat, Sundberg és Skoog (1995) ezek esetében ugyanis nem talált az  $F_1$  emelésére utaló állkapocsnyitási módosításokat, míg Bresch és Narayanan (2010) éppen e hangzóknál mutatta ki legegyszerűbben az  $F_1 : f_0$  hangolást. Az eredmények alapján leszűrhető további fontos és hangsúlyozandó tanulság, hogy az akusztikai adatok azt is mutatják, hogy a szopránénekes által ejteni szándékozott magánhangzók minősége mind a magánhangzó, mind pedig az alapfrekvencia függvényében, az alapfrekvencia emelésével párhuzamosan, megközelítőleg fokozatosan változik. Mindebből következően tehát felvethető (és feltehetőleg egyáltalán nem triviális kérdés), hogy ezek a tipikus megvalósulásokhoz képest többé-kevésbé eltérő, magas alapfrekvencián énekelt magánhangzók miként

sorolhatók be az egyes az adott nyelvre jellemző fonémakategóriák valamelyikébe, azaz milyen hangokként és milyen hatékonysággal azonosíthatók a hallgatók számára.

A fent említett vizsgálatok módszertani bemutatásában a szerzők az alkalmazott módszerekben rejlő hibalehetőségeket is kiemelték, valamint rávilágítottak, hogy az egyes módszerek a megfigyelni kívánt artikulációs-akusztikai stratégiák módosítására készítették az énekeseket. A módszertani problémák közt legfontosabbként az artikulációs adatokkal dolgozó vizsgálatok gyakorlatilag mindegyike a keresztmetszetfüggvény-számításban, valamint a fizikai vagy matematikai modellben végzett rezonanciamérésekben rejlő potenciális mérési pontatlanságokat említette (Bresch és Narayanan 2010 éppen ezért például egyenesen kerülte a konkrét formánsértékek bemutatását és elemzését, csak statisztikailag vetette össze őket az  $f_0$ -értékekkel összehasonlítva). A Rothenberg-maszkot használó akusztikai elemzésen alapuló vizsgálatban a kutatók a maszk akadályozó szerepét emelték ki, ugyanis a maszk alsó pereme, mint írják, korlátozta az énekest az állkapocsnyitás szabad növelésben (Hertegård-Gauffin 1993). Joliveau és munkatársai (2004) valamint Garnier és munkatársai (2010) leírják, hogy az énekest halk és lehetőség szerint teljesen vibrátómentes énekmódra kérték (a mérés technikai megoldás érzékenysége miatt), ami viszont magasabb alapfrekvenciák esetében problematikus lehet. Mivel tudjuk, hogy az  $f_0$  emeléséhez természetesen nagyobb szubglottális nyomás és nagyobb hangerősség járul, feltételezhetjük, hogy az énekhang természetellenes halkítása és a vibrátó nélküli énekprodukciónak némileg befolyásolta a formánsstratégiákat is. Sundberg és munkatársainak (2013) kritikai észrevétele szerint ráadásul a szájnál, fonáció közben bejuttatott külső gerjesztésen alapuló mérés azért is problémás, mert így a fonációs ciklus azon szakaszában, amikor a glottisz nyitva van, nem kívánatos módon a szubglottális üregrendszer rezonátortevékenysége is megjelenik a jelben.

A fent említett vizsgálatok javában azt is láttuk, hogy azok vagy egy hangzóra (tipikusan az /a/-ra) koncentráltak, vagy az artikulációs-akusztikai magánhangzótér szélső értékeit képező /a i u/ hangokra – a legtöbb esetben tehát nem vizsgálták a nyelvspecifikus magánhangzórendszer hatását az egyes magánhangzók megvalósulására. Esetenként, ha a vizsgálatban ezeknél több hangzóminőség szerepelt, a hangzók kiválasztásának oka még mindig nem egy-egy nyelv magánhangzórendszerének a teljes feltérképezése volt, inkább csak az egyes eltérő nyelvállás-



fokok mint képzési jegyek reprezentációja. Az eddigi akusztikai vizsgálatok java nem szentelt tehát különösebb figyelmet a magánhangzók artikulációs-akusztikai változásaiban fellelhető nyelvspecifikus tendenciák megfigyelésére, illetve a teljes magánhangzótér alakulásának sajátosságaira.

Az említett akusztikai (és amint a következő alfejezetben látni fogjuk, a magánhangzók észlelésével foglalkozó percepció) vizsgálatok egy másik méltatlanul mellőzött aspektusa, az artikulációs-akusztikai eredmények és a percepció tendenciák közti összefüggések részletekbe menő elemzése és az akusztikai-artikulációs változásokat követő percepció tendenciák nyelvspecifikus leírása. E mellőzés oka egyfelől persze nagyon is jól érthető, hiszen az alacsony alapfrekvenciájú beszédprodukció percepciójára kidolgozott modellek sem az  $F1 : f_0$  formánshangolás, sem pedig a magas alapfrekvencia okozta alacsony kimeneti felbontás okán nem adnak egyértelmű vagy könnyen igazolható predikciókat. Azonban azt is látnunk kell, hogy egyes kutatók hajlamosak gyakorlatilag teljesen elvitatni a kérdés komplexitását (és sokszor a relevanciáját is), és triviális kérdésnek tekintve azt átsiklanának fölötté egy-két akusztikumból származó egyszerű következtetés után. Joliveau és munkatársai (2004) például azt a predikciót fogalmazzák meg, hogy mivel az alulmintavételezettség miatt már amúgy is egyértelműen csökken a magánhangzó érthetősége magas alapfrekvencián, nem meglepő, hogy az énekesek a hangzóminőségek megtartása helyett a hangerősség és hangszín produkciójára koncentrálnak a formánsok értékének változtatását (az  $F1 : f_0$  hangolást) választják stratégiaként. Értelmezésük szerint tehát a magánhangzók érthetőségének csökkenése már pusztán a spektrális alul-mintavételezettség fokozódásából következően is szükségszerű (és egyértelmű), amelyen az  $F1$  hangolása már „nem ront sokat”, és a két faktor együttes hatása gyakorlatilag felismerhetetlenné teszi a magánhangzókat.

Hangsúlyozni kívánjuk, hogy a rezonanciaadatok és a percepció adatok összefüggésbe hozásával kapcsolatos aggályokat és elvi-gyakorlati nehézségeket ismerjük, és érthetőnek tartjuk, hiszen a magánhangzó-észlelés hagyományos, formánsokon és azok viszonyán alapuló értelmezése a magas alapfrekvenciák esetében valóban nehezen értelmezhető, vagy egyenesen akár értelmezhetetlen is lehet. Ugyanakkor úgy véljük, hogy nem lenne haszon nélkül való, ha szisztematikus áttekintést nyerhetnénk az akusztikai és percepció adatok közti lehetséges össz-

szefüggésekről egy nyelv (lehetőség szerint) teljes magánhangzókészletének vizsgálata segítségével – és persze minden, a magas alapfrekvenciából és a formánshangolás jelenségéből következő megszorítást szem előtt tartva.

A következőkben a magas alapfrekvencián énekelt magánhangzók észlelésének kérdésében eddig folytatott vizsgálatok eredményeit ismeretjük.

### **1.2.3. Az énekelt magánhangzók percepciója**

Az énekelve ejtett magánhangzók eltérése a mindennapi beszédben megvalósuló magánhangzókhoz képest jóval nagyobb lehet, mint a magánhangzókra a hangkörnyezettől és a beszélőtől függő variabilitás következtében jellemző változatosság. Az éneklés ilyen szempontból tehát a beszéd határhelyzetének, extremitásának tekinthető – úgy értelmezhető, mint egy beszédprodukciós mód, mely egyfajta „természetes” módon feszegeti a beszédhangok megvalósításának határait. A különleges produkciós mód révén egyúttal a beszédpercepció is új oldaláról válik vizsgálhatóvá, ha ezen extrémnek tűnő beszédhang-megvalósítások észleleti vetületét elemezzük.

Noha úgy tekinthetjük, hogy az éneklés például a laboratóriumi körülmények között felvett és akár gépileg manipulált beszéddel szemben a beszédprodukció valamely természetes extremitását képezi, nélkülözi a laboratóriumi beszéd vizsgálatának egyes nagyon hasznos sajátosságait is. Az énekelt magánhangzók vizsgálatában ugyanis bizonyos jelenségek – legfőképpen az alapfrekvencia és a formánsfrekvenciák alapfrekvenciafüggő változásai – nem választhatók szét egymástól, hiszen az éneklésben nincsen egyik jelenség a másik nélkül, nem volna tehát az éneklésre értelmezhető egy a jelenségek disszociációján alapuló kísérlet. De ragadjuk meg a kérdést mégis inkább a másik oldaláról, és máris optimistább képet kapunk az énekelt beszédhangok vizsgálatának hasznosságáról. Éppen abban rejlik ugyanis az énekelt magánhangzók érdekessége, hogy azok egyes olyan jelenségeknek a természetes ötvözetét képezik a mindennapokban tapasztalható beszédhang-megvalósításban, melyekről a percepciós szakirodalom alapján tudjuk, hogy befolyásolják, sőt alapvetően meghatározzák az egyes magánhangzó-minőségeket, azok észlelését. A jelenségek az éneklésben tapasztalható komplex érvényesülésének észleleti hatása azonban egyáltalán nem magától értetődő (nem válaszolható

meg pusztán a már ismert beszédpercepciós modellek valamelyikének egyszerű adaptációjával), a kérdés megválaszolásához kísérleti bizonyítékok szükségesek.

A következőkben előbb ismertetjük, milyen predikciókat fogalmaznak meg az egyes az akusztikummal foglalkozó tanulmányok az énekelt magánhangzók azonosítására és azonosíthatóságára, utóbb pedig bemutatjuk a kérdés eddigi kísérletes eredményeit.

#### **1.2.3.1. Az énekelt magánhangzók észlelésével kapcsolatos gyakori feltételezések**

Az énekelt magánhangzók artikulációs és akusztikai szerkezetének ismeretében számos predikció megfogalmazható az énekelt magánhangzók észlelésére. Látni fogjuk azonban, hogy ezen feltételezéseknek csak egy elenyésző részével kapcsolatban áll rendelkezésünkre megbízható kísérletes bizonyíték. Amint azt például Joliveau és munkatársai (2004) is megfogalmazzák, „biztosra vehető”, hogy a magánhangzók magas alapfrekvenciákon énekelve az alapfrekvencia emelésével minőségükben egymáshoz egyre hasonlóbba lesznek, míg végül kellően magas alapfrekvencián teljesen uniformizálódnak, ráadásul úgy, hogy az  $f_0$  emelését kísérő egyre növekvő állkapocsnyitásszöggel és egyre magasabb  $F_1$ -gyel ejtett magánhangzók egyre nyíltabb magánhangzókként lesznek felismerhetők. A szerzők ezt a felvetést annak ellenére fogalmazzák meg, hogy a magánhangzók nyíltságának észleletéről már több vizsgálat is bizonyította, hogy nem elsősorban nem az  $F_1$  abszolút értékének, hanem annak az  $f_0$ -hoz viszonyított tonotopikus távolságának függvénye (vö. 1.1.2.2. *alfejezet*). Egy ehhez toldott további gyakori feltételezés továbbá, hogy a mássalhangzó-környezet a koratikulációs hatásokon keresztül minden bizonnyal „javít” az ejtési szándék szerinti magánhangzó-azonosításon még a magas alapfrekvencián énekelt magánhangzók esetében is (vö. pl. Scotto di Carlo–Germain 1985, Gottfried–Chew 1986, Hollien et al. 2000). Az ezt az állítást megfogalmazó szerzők azonban az esetek többségében nem foglalnak egyértelműen állást abban a kérdésben, hogy a mássalhangzók hatását a percepciós vagy az artikulációs-akusztikai vetületben feltételezik-e (erről még az egyes percepciós tanulmányok tárgyalásánál több szót ejtünk).

Amint ezt a következő alfejezetben látni fogjuk, mindkét feltételezés esetében igaz, hogy inkább csak spekulatív, de nem szigorú kísérletekkel ellenőrzött feltevéseken alapszik. A spe-

cifikusan e kérdések vizsgálatát célzó, nagyon szisztematikus és jól kontrollált kísérletek ugyanis a mai napig hiányoznak a szakirodalomból.

#### 1.2.3.2. Az énekelt magánhangzók észlelésének vizsgálata

Sundberg 1975-ben született tanulmánya óta kísérletes bizonyítékokkal rendelkezünk arról, hogy az énekelt magánhangzók akusztikuma az alapfrekvencia és a magánhangzó-minőség függvényében többé vagy kevésbé eltér a beszédben ejtett magánhangzók akusztikai szerkezetétől. Amint azt az előző alfejezetben bemutattuk, Sundberg (1975) és az őt követő szerzők artikulációs és akusztikai elemzéseinek értelmében a magas alapfrekvencián énekelt magánhangzók és beszédbeli megvalósulásuk közti legfőbb különbség az első formánsnak az alapfrekvencia értékére (vagy valamivel a fölé) hangolása azokban az esetekben, amikor a megemelt alapfrekvencia magasabb lenne a magánhangzóra a beszédben jellemző első formáns frekvenciaértékénél. Az előző alfejezetben ezen túlmenően azt is kiemeltük, hogy a beszéd és a magas alapfrekvenciájú éneklés között további jelentős eltérések származnak a magas alapfrekvencia okozta spektrális alul-mintavételezettség jelenségéből is, azaz abból a pusztán alapfrekvencia-függő jelenségből, melynek eredményeképpen a magas alapfrekvenciákon ejtett magánhangzók felbontása alacsonyabb az akusztikai kimenetben az alacsonyabb alapfrekvenciákon ejtett magánhangzókhoz képest.

Mindezen felfedezések eredményeképpen az idézett Sundberg-tanulmány (1975) megjelenése óta az is egyre terjedő nézetnek számít a szakirodalomban, hogy a magas alapfrekvenciájú (szoprán) éneklésben ejtett magánhangzók „érthetősége” (az angol *intelligibility* szóból fordítva) vagy észlelhetősége, azaz e magánhangzók az ejtési szándék szerinti azonosítása is alacsony. Valamivel Sundberg (1975) vizsgálata után ezt a sok esetben kísérletes igazolás nélkül elfogadott feltételezést több vizsgálat is górcső alá vette. E vizsgálatok központi kérdéseit a magas alapfrekvencián énekelt magánhangzókra az alapfrekvencia függvényében jellemző azonosítási tendenciák, valamint az azonosítást segítő vagy éppen akadályozó körülmények (pl. a fonetikai kontextus vagy annak hiánya) képezték. Az elkövetkezőkben ezekről a vizsgáltokról lesz szó.

Mielőtt azonban rátérnénk az énekelt magánhangzók kísérletes vizsgálatának eddigi eredményire, fontosnak tartjuk ehelyütt egy, a jelen értekezésben az énekelt magánhangzók ejtésével és észlelésével kapcsolatban elfogadottnak tekintett elképzelések szempontjából alapvető fogalmi kérdés tisztázását. A nemzetközi (angol nyelvű), az énekelt magánhangzók azonosításával foglalkozó szakirodalomban az eredmények értékelésénél általánosan azt találjuk, hogy a válaszként megjelent magánhangzókat az ejteni szándékozott magánhangzóhoz viszonyítva „helyes” (*correct*) vagy „hibás/téves” (*incorrect/mistake/error*) válaszokként minősítik a szerzők. Teszik ezt azonban annak ellenére, hogy a legtöbb esetben az adott vizsgálatban nem térnek ki az ejtett hangzók akusztikai szerkezetének elemzésére, pusztán az ejteni szándékozott magánhangzókat listázzák, és ezekhez viszonyítják a válaszokat. A „helyes válasz” és „téves válasz” megfogalmazások tehát erősen félrevezetők, hiszen azt sugallják, hogy a válaszok kiértékelésekor pontosan tisztában vagyunk az adott beszédhangoknak az akusztikai szerkezet által definált minőségével, azaz a „helyes válasszal”. Mindez azonban mind Sundberg (1975) eredményeinek, mind pedig minden későbbi, az énekelt magánhangzók akusztikai szerkezetét vizsgáló eredménynek ellentmond (vö. előző alfejezet), hiszen ezek a vizsgálatok éppen a magánhangzók a hangzóminőség és az alaphangfrekvencia függvényében az alaphangfrekvencia növelésével egyre inkább megváltozó akusztikai szerkezetéről tanúskodtak.

Az előző alfejezetben ismertetett akusztikai vizsgálatok egyik legfontosabb – de a szakirodalomban talán nem kellően hangsúlyozott – tanulságának éppen azt tartjuk, hogy azok eredményei szerint a magánhangzók realizációja az éneklésben az adott magánhangzó minősége és az alaphangfrekvencia függvényében az alaphangfrekvencia emelésével mind a hangzók spektrális mintavételezettsége, mint pedig a hangzók formánsszerkezete tekintetében megközelítőleg fokozatosan változik. Ennek értelmében tehát a magas alaphangfrekvenciákon énekelt magánhangzók minősége, a magánhangzó-minőségek fonemikus besorolása nem triviális, sőt vélekedésünk szerint a fonemikus besorolást még akusztikai adatok birtokában sem tekinthetjük egyértelműnek vagy adottnak. A fent említett változások ugyanis – az eddigi eredmények alapján – alapvetően nem kategóriaváltás-szerűek, azaz nem egy, az adott nyelven létező beszédhang egy másik, az adott nyelvben létező beszédhangként való megvalósításaként írhatók le, hanem graduálisak. Ez viszont azt jelenti, hogy a megváltozott akusztikai szerkezetű hangok percep-

ciós vetülete éppen a beszédpercepció természetéből fakadóan nem lehet triviális, és e módosult akusztikai szerkezetű hangok besorolása csak kísérletes úton állapítható meg.

A fentiek értelmében úgy gondoljuk, hogy az énekelt magánhangzók percepció (és ezzel kiegészülő akusztikai) vizsgálatának egyik legfontosabb célja az alaphfrekvencia változásának következtében módosult akusztikai szerkezetű hangok percepció vetületének elemzése. A tesztekben kapott válaszok „helyes” és „téves” minősítésével tehát éppen a vizsgálatban megcélzott kérdést „válaszolnánk meg” előre (ráadásul sok esetben valószínűleg tévesen). Ennek megfelelően az értekezésben lehetőség szerint kerüljük a magánhangzók azonosításának „helyes” vagy „téves” minősítéseit, helyette a magánhangzók az ejtési szándék szerinti vagy attól eltérő (fonemikus) besorolásáról, a magánhangzók azonosításáról beszélünk, így kerülve el azt a véleményünk szerint értelmezhetetlen állásfoglalást, hogy a kísérletekben részt vevő hallgatóknak milyen magánhangzót kellett volna felismerniük a kategóriahatárokat feszegető, nem egyértelmű akusztikai szerkezetű hangokban. Vegyük észre azonban, hogy még valamilyest az általunk használt megjelölés is félrevezető lehet, ha arra gondolunk, hogy az ejtési szándék szigorúan véve nem csak az adott feladatban az énekes előtt szereplő fonémalista megvalósításának szándékát, de már az adott esetben, az adott alaphfrekvencián esetleg szükséges formánshangolás (vagy más, az éneklés során elsajátított formánsstratégiák) alkalmazásának intencióját is magában foglalja. Zavarónak tartanánk azonban egy még ennél is pontosabb, így szükségszerűen komplexebb, terjedősebb megfogalmazás használatát az értekezésben. Éppen azért megmaradunk az „ejtési szándék szerinti azonosítás” formulánál, azzal a kitétel-lel, hogy az „ejtési szándékot” kizárólag az énekes adatközlőknek a felvételekkor mutatott stimulus-, vagy fonémalista által megjelölt magánhangzó-minőségek megvalósításának szándékára értjük, nem pedig az egyes magánhangzók az énektechnika elsajátítása során tanult, az alaphfrekvenciától függően módosított megvalósításának stratégiájára.

A következőkben az énekelt magánhangzók azonosítására vonatkozó szakirodalom legtöbbet hivatkozott, azaz legnagyobb befolyásúnak tűnő tanulmányait ismertetjük.

Smith és Scott (1980) azt a kérdést igyekezett körüljárni, hogy lehetséges-e – és ha igen, miként – a magas alaphfrekvencián énekelt magánhangzókat érintő akusztikai változásoknak (de főként az  $F1 : f0$  formánshangolásnak) a kiküszöbölése az észleletben. Kísérletükben a

szerzők abból az alapfeltevésből indultak ki, hogy a magas alapfrekvencián énekelt magánhangzók akusztikai szerkezete megváltozik, és ebből egyenesen következik az is, hogy az ejteni szándékozott magánhangzó az ejtési szándék szerinti azonosításának aránya egyre csökken az alapfrekvencia emelésével. Az  $F1 : f_0$  formánshangolásról bevezetjük értelmében azt feltételezték, hogy elsősorban az alacsony első formánssal rendelkező, azaz a zárt, első nyitásfokú, felső nyelvéllású magánhangzók éneklésekor fellépő jelenség, így elsősorban e hangzók esetében jelenthet az észleleti magánhangzó-minőséget érintő problémát. Smith és Scott (1980) Strange és munkatársainak (1976) kísérleti eredményire építve azt a feltételezést igyekezett megvizsgálni, mely szerint a magánhangzók az ejtési szándék szerinti azonosítása hatékonyabb, ha a magánhangzót mássalhangzó-környezetben ejtve halljuk. (Strange és munkatársai 1976 vizsgálataikban azt igazolták, hogy a magánhangzót körülvevő mássalhangzók a magánhangzó és a mássalhangzók közötti hangátmenetekben rejlő akusztikai információk keresztül segítik a magánhangzó felismerését akár értelmetlen szavakban is, sőt még akár olyan helyzetekben is, amikor a teljes magánhangzós rész törlődött a CVC szekvenciából, vö. a jelen értekezés *1.1.2.2. alfejezetét*).

Mivel az énektanítás elméletének történetében már nagyon régóta létező feltételezés (vö. Garcia 1911) – valamint a Smith és Scott (1980) vizsgálatát megelőző egyes kísérletes eredmények – szerint a jó szoprán énektechnika a gége lesüllyesztését alkalmazza (vö. a jelen értekezés *1.2.1 alfejezetét*), a szerzők a mássalhangzó-környezet hatásán kívül a gégehelyzet hatását is tesztelni kívánták percepciós vizsgálatukban.

A kísérlet anyagát az angol /i ɪ ε æ/ magánhangzók képezték a *bVd* hangkörnyezetben (azaz a *bead* 'gyöngy', *bid* 'ajánlat, tét', *bed* 'ágy', *bad* 'rossz' minimális négyesben, „CVC” kondíció), és izolációban („V” kondíció), az *f'* (349 Hz), *a'* (440 Hz), *c#''* (554 Hz), *f''* (698 Hz), *a''* (880 Hz) és *c#'''* (1108 Hz) alapfrekvencián, egy szoprán énekes énekprodukción. A „V” kondícióba tartozó magánhangzókat a szerzők izolált ejtésben vizsgálták, azaz graduális zöngéindítást, hangindítást vagy zöngékezdetet (onset) és graduális zöngелеcsengést (offset) tartalmazó hangokban. A két, a gégehelyzettel összefüggő kondíciót a szerzők „operai” és „megemelt gége” énekprodukciónak nevezték, és bár ezek elicitálásának módjáról és kontrollálásáról nem adnak pontosabb leírást, a módozatokat valamelyest definiálták tanulmányuk-

ban: meghatározásuk szerint az „operai” produkciós mód (a „megemelt gégevel” szemben) a „rendes” operaénekülésben használt technika. A percepció tesztet 10 résztvevővel (egyetemi hallgatóval) végezték el, a szerzők az egyes magánhangzókat randomizált, de az egyes al-frekvenciákat frekvenciájuk szerint egymást követő, növekvő sorrendben játszották le a kísérlet során.

Smith és Scott eredményei alapján a legmagasabb (1108 Hz) al-frekvencián az ejtési szándék megfelelő azonosítás szerint az egyes kondíciók közt tapasztalt sorrend a következő volt: „megemelt gége” + „CVC” (76%) > „operai” + „CVC” (70%) > „megemelt gége” + „V” (44%) > „operai” + „V” (12%). Alacsonyabb frekvenciákon a kondíciók itt bemutatott sorrendje kevésbé mutatkozott – mert például a „megemelt gége” + „V” kombináció sok esetben jóval, akár 32%-kal is magasabb százalékos arányokat mutatott, mint az „operai” + „CVC” kombináció. Az „operai” + „V” kombináció minden al-frekvencián elmaradt a többi kombinációhoz képest, míg a „megemelt gége” + „CVC” kombináció gyakorlatilag minden al-frekvencián a legmagasabb azonosítási százalékokat mutatta (azaz itt egyezett a legnagyobb számban az ejtési szándék szerinti magánhangzó-minőség az azonosított magánhangzóval). Az „operai” mód a „V” és „CVC” kondícióban kapott eredményei a magas ( $f''$  vagy 698 Hz-es és afeletti) al-frekvenciákon szisztematikusak voltak, és a mássalhangzó-környezetben ejtett hangok előnyét mutatták (rendre 16 és 54%, 4 és 68%, valamint 12 és 70%). Smith és Scott a  $f''$  (698 Hz) feletti, tehát az általuk magasnak tekintett al-frekvenciák kiátlagolásával a következő sorrendet állította fel a kondíciópárok között: „megemelt gége” + „CVC” (83%) > „operai” + „CVC” (64%) > „megemelt gége” + „V” (62%) > „operai” + „V” (10%).

Eredményeiket a szerzők a következőképpen foglalták össze. 1. A magas al-frekvencián operai módban énekelt, izolált ejtésű magánhangzók azonosítása alacsony – ezt a szerzők korábbi vizsgálatok eredményeivel egyezőnek tartották, és további magyarázatra nem szoruló megfigyelésként jegyezték. 2. Az izolált ejtésű magánhangzók azonosítása magasabb a „nem rendes” ének-módban, mint a „rendes”, operai módban. Ennek magyarázataként – bár arra egyértelmű akusztikai bizonyítékot nem találtak – a szerzők feltételezték az  $F1 : f_0$  formánshangolás megjelenését az operai ének-módban. Magyarázatuk szerint tehát a kettes számmal jelölt eredményekkel kapcsolatban a két kondíció közti különbséget a hangolási stratégia megléte (a



„rendes” operai produkciós módban) vagy hiánya (a „nem rendes” produkciós módban) okozhatta, hiszen a hangolás eltérést okoz a magánhangzókban a beszédben jellemző akusztikai szerkezetéhez képest. 3. A megemelt gégével képzett („nem rendes”) énekhangban az izolált ejtésű magánhangzók azonosítási aránya alacsonyabb volt, mint az azonos módban, de mássalhangzó-környezetben ejtett magánhangzók azonosítási aránya. A szerzők magyarázata szerint ennek legfőbb oka az lehetett, hogy a mássalhangzó-környezet a formánsátmenetek révén segítette a magánhangzók percepcióját még magas alapfrekvenciájú hangok esetében is. 4. Az operai technikával képzett énekhangban a mássalhangzó-környezetben ejtett magánhangzók azonosítása magasabb volt, mint az azonos módban képzett énekhang izolált magánhangzóiban, még magas alapfrekvenciákon is. A szerzők itt magyarázatként ismét csak azt sugallják (de óatosan kerülnek a konklúzió explicit levonását), hogy ez az eredmény a mássalhangzó-környezetnek, valamint a formánsátmeneteknek köszönhető, tehát a mássalhangzó-környezet még magas alapfrekvenciájú hangok esetében is segíti a magánhangzók az ejtési szándék szerinti azonosítását.

Fontos észrevennünk, hogy a szerzők utóbbi érvüket annak ellenére javasolják, hogy a magasabb alapfrekvenciájú, operai ejtésű magánhangzók esetében maguk is formánshangolást, tehát a magánhangzónak az alapfrekvenciával változó akusztikai megvalósulásait feltételezték. Korábbi érvük értelmében a szerzők feltételezése szerint az  $F1 : f0$  formánshangolás mint az operai énekmod sajátja módosítja a beszédben jellemző akusztikai szerkezetet, és így csökkenti az ejtési szándék szerinti azonosíthatóságot – ezzel magyarázható, hogy az izolált ejtésű magánhangzók azonosítása alacsonyabb volt az operai módban a „nem rendes” énekprodukcióhoz képest. Ha viszont ennek ellenére a mássalhangzó-környezetben mégis magasabb a magánhangzók ejtés szerinti azonosítása, mint izolált ejtésben, a mássalhangzók hatására vonatkozó magyarázat nem lehet a magánhangzók észleleti vetületére vonatkoztatható megfigyelés. A mássalhangzó-környezet pozitív hatásának feltevése ugyanis azt jelenti, hogy a magánhangzókat az alapfrekvencia emelésével érintő akusztikai (és az emögött meghúzódo artikulációs) változások, azaz az  $F1 : f0$  hangolás a kontextus függvényében érvényesült. A negyedik pontban összefoglalt eredmény ugyanis a formánshangolás feltételezése mellett csak kétféleképpen értelmezhető: 1. a formánshangolás csak a kontextussal nem rendelkező magánhangzókat érin-

tette, vagy 2. a formánshangolás a konextussal rendelkező magánhangzókat az izolált magánhangzóknál kevésbé érintette. Amennyiben tehát a szerzők azt sugallják, hogy a mássalhangzó-környezet még a magas alapfrekvenciájú operai módban is segíti a magánhangzóknak az ejtési szándék szerinti azonosítását, az azt is impikálja, hogy a szerzők eredményei és következtetései nincsenek teljes összhangban a kiinduló feltételezésekkel, hiszen érveik és magyarázataik ilyen módon nem a magánhangzók kontextusfüggő azonosítására, hanem azok kontextusfüggő megvalósítására (artikulációs és/vagy akusztikai sajátosságaira) vonatkoz(ná)nak.

Fontos megfigyelnünk azt is, hogy Smith és Scott (1980) tanulmánya nem közöl sokat a percepciós tesztek pontos körülményeiről – így nem tudjuk meg például, hogy az adatközlők milyen instrukciók mellett szolgáltatták válaszaikat. Annyi azonban bizonyos, hogy a válaszadás verbálisan történt – azaz a résztvevők szóban jelezték a kísérletvezetőnek, mely hangot vélték hallani –, valamint az is tudható, hogy a kísérletben használt CVC szekvenciák értelmes szavak voltak. A válaszadás modalitásából feltételezhetjük, hogy a kísérletvezető egyik kondícióban sem prezentált válaszlehetőségeket a résztvevőknek. Ha pedig nincsenek előre ismertett válaszlehetőségek, valószínűsíthető az is, hogy a kísérlet során más módon sem adtak megszorításokat a kutatók a hallgatónak a válaszlehetőségek számának vagy minőségének tekintetében. Ha pedig így van, az azt jelenti, hogy míg a CVC kondícióban négy<sup>13</sup> lehetőség-ből, négy értelmes szó közül választhattak a hallgatók, addig az izolált ejtésű hangok esetében a választási lehetőségek egyáltalán nem korlátozódtak arra négy hangzóra, amit az értelmes értelmes szavak megengedtek – az izolált ejtésben ugyanis a válaszlehetőségek számát csak az angol nyelv magánhangzókészlete korlátozta, mely legalább tíz<sup>14</sup> lehetséges magánhangzóválaszt enged meg. Ennek fényében Smith és Scott (1980) a magánhangzóknak a mássalhangzó-környezet meglététől függő azonosítását valójában két jelentősen eltérő percepciós működést

---

<sup>13</sup> A választási lehetőségek száma elméletben ötre nő, ha feltételezzük, hogy az instruktor nem ismeretette a szólistát a kísérlet elején, mert a fent listázottakon kívül szóbajöhet még a *bayed* szó is, azaz az /eɪ/ diftongus mint válaszlehetőség. Ugyanakkor az is felelhető, hogy a kitartott (azaz a monoftongusejtésnek kedvező) énekelt magánhangzók esetében ennek a hangzónak az előfordulása az észleletben nagyon alacsony valószínűségű, tehát ez a beszédhang jó eséllyel nem, vagy csak kisszámban merülne fel az ejtési szándéktól eltérő azonosítások esetén.

<sup>14</sup> Csak az angol monoftongusokat számba véve ugyanis izolált ejtésben az /i ɪ e ε æ ʌ o u/ hangzók jöhetnek szóba válaszként.

összevetve, két jelentősen eltérő kondícióban igyekezett vizsgálni. Az egyik kondícióban az értelmes szavak révén a nyelv magasabb szintjei, a morfológia, illetve a szemantika is befolyásolta az azonosítás tendenciáit, hiszen ezek korlátozták a válaszlehetőségeket és a válaszlehetőségek számát a válaszként szóba jöhető létező szavak (azaz a top–down percepciós folyamatok) révén. Ám a másik, az izolált ejtésű hangzókat tartalmazó kondícióban ezzel szemben csak az alacsonyabb szintű, elemi hallási észlelésen és az akusztikai elemzésen alapuló (és csak az akusztikai kulcsokra támaszkodó) beszédpercepciós (bottom–up) folyamatok léphettek működésbe úgy, hogy a lehetséges válaszok számát az angol nyelv magánhangzókészletén kívül semmi sem korlátozta. Nem állítható tehát bizonyosan, hogy a Smith és Scott (1980) vizsgálatában kapott eredmények nem pusztán az egyes kondíciókban lehetséges válaszok számának, azaz az egyes válaszok előfordulási valószínűségének, valamint a helyes tippelési esély kondícióként eltérő mértékének következményei, hanem valóban a mássalhangzó-környezet pozitív hatását mutatják.

Végezetül pedig megemlíteném az is, hogy a Smith és Scott (1980) által vizsgált négy magánhangzó minősége is erősen befolyásolhatta a kapott eredményeket. A közelmúlt egyes, más kérdéseket megválaszolni kívánó, de a jelen problematika szempontjából érvényes eredményekkel szolgáló vizsgálatai alapján ma már tudható, hogy a Smith és Scott (1980) által vizsgált szavak és magánhangzók (egy kivételével)<sup>15</sup> a gondolozott, túlartikulált vagy hiperartikulált, lassú beszédben (a szerzők megfogalmazásában *clear speech*-ben) nagyobb számban ismerhetők fel, mint a hétköznapi, normál beszédben (Rogers et al. 2010). Mivel pedig az éneklési feladatban Smith és Scott vizsgálatában az énekes hosszú, azaz egyúttal feszebb, „túlartikuláltabb” magánhangzókat ejtett, feltehető, hogy azon alaphangfrekvenciák esetében, amelyekben még az egyes magánhangzóknál nem lépett fel artikulációs-akusztikai változtatás (F1 : f0 hangolás), az énekelt magánhangzók az ejtési szándék szerinti azonosítása a beszédhez hasonlóan jó arányú lehetett. Ezen túlmenően pedig azt is észre kell vennünk, hogy a Smith és Scott (1980) által vizsgált négy beszédhangból kettő, mégpedig az /ɛ/ és az /æ/ nyíltabb (harmadik és negyedik nyitásfokú) hangzók – melyekről nyíltsági fokuk és magas F1 ér-

---

<sup>15</sup> A kivételt a *lax*, azaz laza ejtésű /ɪ/ képezte, mely esetében a gondozott vagy túlartikulált beszédben a normál beszédejtéshez képest alacsonyabb volt a magánhangzók azonosításának aránya (Rogers et al. 2010).

tékük miatt feltehető, hogy még magas alapfrekvencián is jelentősebb változtatás (azaz  $F1 : f_0$  hangolás) nélkül voltak ejthetők a kísérletben résztvevő énekes számára. Úgy tűnik tehát, hogy a magas alapfrekvencián is jelentősebb változtatás nélkül ejthető és a „túllartikuláció” miatt feltehetően jól azonosítható hangok Smith és Scott (1980) vizsgálatában a „CVC” kondíció tesztanyagának pontosan felét képezték, ugyanakkor az egyes magánhangzók azonosítási tendenciáit csak összevonva ismertetik a szerzők. Ebből következően a magánhangzók minőségbeli jellemzőiből fakadó pozitív és negatív hatások összevegyülnek és kioltják egymást a vizsgálatban, így módon valamelyest feltehetően félrevezető képet festve a tapasztalt észlelési tendenciákról.

Noha Smith és Scott (1980) konklúziói alapján úgy tűnik, hogy a vizsgálat a mássalhangzó-környezet és a koartikulációs formánsátmenetek pozitív hatását mutatta ki a magas alapfrekvencián énekelt magánhangzóknak az ejtési szándék szerinti azonosítására, a fentebb ismertetett körülmények (a vizsgált magánhangzóknak és a kondíciók eltérő feltételeinek) fényében az is valószínűsíthetjük, hogy a vizsgálati dizájn *confound* (azaz több, egymással interakcióban lévő) hatás révén támogatta, erősítette a különbségeket a mássalhangzó-környezet hatását reprezentálni kívánó kondíció és az izolált ejtést reprezentáló kondíció között. Ennek fényében a mássalhangzó-környezet pozitív hatására vonatkozó következtetést óvatosan kezelendőnek és további vizsgálatra érdemes kérdésnek tartjuk.

Gottfried és Chew (1986) az angol /i ɪ e ε æ ʌ o u ʊ/ magánhangzókat vizsgálta férfi éneklésben, mégpedig egy kontratenor énekprodukción. Észlelési tesztjeikkel a szerzők arra keresték a választ, hogy miként változik a magánhangzók azonosítása az  $f_0$  és a fonációs mód vagy (zöng)regiszter függvényében (az utóbbi kérdést a mondális zöng és az általában a férfiak által a magasabb alapfrekvenciákra használt falzett szembenállásával reprezentálva), valamint a mássalhangzó-környezet hatására. Ehhez a *bVd* hangkörnyezetben (azaz tíz létező angol szóban) rögzítették az imént felsorolt magánhangzókat mindkét fonációs módban: mondális zöngével az *c* (130 Hz), *e* (165 Hz), *a* (220 Hz), *c'* (260 Hz) és *e'* (330 Hz), falzett fonációval pedig az *a* (220 Hz), *c'* (260 Hz), *e'* (330 Hz), *a'* (440 Hz), *c''* (520 Hz) zenei hangokon (a kétféle fonációs móddal produkált alapfrekvencia-tartományok tehát átfedésben voltak). Ezután a kutatók kivágták minden szótag középső, megközelítőleg 200 ms hosszúságú

kvázi-stacioner részét (azaz a magánhangzó „tisztá fázisát”, vö. Gósy 2004), így létrehozva egy, a mássalhangzó-környezetben ejtett magánhangzókkal, azaz a „teljes szótag” kondícióval opponáló „magánhangzó” kondíciót. A vizsgálat során a szerzők összesen négy percepciós tesztet végeztek el. Ezek során rendre a következő feltételpárokat tesztelték: „modális” + „teljes szótag”, „modális” + „magánhangzó”, „falzett” + „teljes szótag”, „falzett” + „magánhangzó”. Az egyes opponáló feltételeket („teljes szótag” és „magánhangzó”, valamint „modális” és „falzett”) tehát külön-külön vizsgálták.

A tanulmányban ismertetett eredmények szerint mind az  $f_0$ , mind pedig a fonációs mód hatással volt a magánhangzók az ejtési szándék szerinti azonosításának százalékos arányára: az ejtési szándéktól eltérő válaszok aránya az  $f_0$  emelésével nőtt, ám a modális regiszter felső tartományában talált nagy „hibaszázalékot” a fonációs mód váltása, azaz a falzett képzésmódra való áttérés csökkentette, míg az  $f_0$  további emelésével az ejtési szándéktól eltérő válaszok aránya ismét nőni kezdett. Ezen túlmenően a vizsgálat azt is kimutatta, hogy a veláris hangzókra átlagosan nagyobb arányban érkezett az ejtési szándéktól eltérő válasz, valamint hogy a teljes szótagra minden alapfrekvencián nagyobb arányban érkeztek az ejtési szándéknak megfelelő válaszok, mint a magánhangzók tiszta fázisát tartalmazó stimulusokra. Az akusztikai elemzések (LPC-analízis, melynek eredményei a magas alapfrekvencián tapasztalható spektrális alul-mintavételezés figyelembevételével értelmezendők) szerint a regiszterek, azaz a kétféle zöngképzési mód közti különbségek az F1 és F2 formáns értékére is hatással voltak, mégpedig a következőképpen. A beszédben alacsonyabb F1 formánssal rendelkező első, második vagy harmadik nyitásfokú hangzók esetében a falzett módban a beszédértékhez közelebbi F1 volt mérhető, míg ugyanezen fonációs módban a zárt és veláris hangzókban a beszédben mérhetőhöz közelebbi frekvencián megvalósuló F2-t találtak. (Megjegyzendő, hogy nem teljesen egyértelmű, hogy a szerzők az alacsonyabb F1- és F2- értékeket nem az LPC-analízis téves formánsdetekciójaként értelmezik-e.) Gottfried és Chew (1986) következtetései szerint a magasabb alapfrekvencián ejtett magánhangzókra kapott alacsonyabb azonosítási arányok oka egyaránt (és akár egymástól függetlenül) lehet a magánhangzók az alacsonyabb alapfrekvenciákon tapasztalttól eltérő artikulációja, valamint a magas alapfrekvencián a spektrumban megjelenő felharmonikusok kisebb száma is (azaz a magánhangzók alacsony felbontása vagy

másként a magas alapfrekvencián ejtett szonoránsokat jellemző spektrális alul-mintavételezettség, vö. *1.1.1. alfejezet*).

Cikkük bevezetőjében Gottfried és Chew (1986) hivatkozik azokra az egyes hangszerek hangjának felismerését vizsgáló korábbi pszichoakusztikai eredményekre, melyek szerint egy adott alapfrekvencián megszólaló hang hangszínének azonosításában, azaz a hangszerek keltette hangok világában a hangot kiadó hangszer felismerésében kiemelt szerepű akusztikai kulcs a hangindítás; az idézett tanulmányok értelmében ugyanis a hangindítás eliminálásával jelentős mértékben csökken a hangszerek felismerésének sikeressége (pl. Berger 1964, Saldanha–Corso 1964). Bár az említett eredmények citálása azt a benyomást kelti, hogy a szerzők elfogadják azt a feltételezést, mely szerint a csak a hangszínükben eltérő hangok megkülönböztetésében a hangadás kezdetének kiemelt jelentősége van, valamint úgy tűnik, hogy a feltételezést a beszédhangokra (tehát két különböző minőségű magánhangzó megkülönböztetésére nézve) is kiterjeszthetőnek tartják, érdekes – és egyben igen sajnálatos – módon mégsem ellenőrzik azt vizsgálatukban. Az izoláltan ejtett hangok helyett ugyanis csak a szótagmagból kivágott kvázi-stacioner magánhangzó-szeleteket tesztelik. Mivel pedig ily módon a tesztben izoláltan lejátszott magánhangzók nem valódi izoláltan ejtett magánhangzók voltak (melyek ennél fogva az izolált magánhangzók karakterisztikus hangindítási sajátosságaival sem rendelkeznek), a vizsgálat eredményei nem vonatkoztathatóak egyértelműen a valódi izoláltan ejtett, illetve a mássalhangzó-környezetben ejtett magánhangzók azonosításának különbségeire. Ezt az egyébként releváns és érdekesnek ígérkező összehasonlítást tehát Gottfried és Chew (1986) bár implicit felvetni látszik, mégis nyitott kérdésként hagyja.

Gottfried és Chew (1986) egyik javaslata szerint tehát a magánhangzók a kísérletben tapasztalt, az alapfrekvencia emelésével összefüggően növekvő ambiguitása nem feltétlenül az f<sub>0</sub> emelését kísérő artikulációs és az azzal járó akusztikai változások, hanem a spektrális alul-mintavételezettség következménye is lehet. Amennyiben pedig ez a feltételezés igaz, és a magánhangzókat tényleg nem jellemzi más – jelentősebb – változás, mint a fokozódó spektrális alul-mintavételezettség, jól beláthatónak tűnik a szerzők azon magyarázata is, mely szerint (a kísérletben vizsgált 520 Hz alatt) a mássalhangzó-környezet a koartikulációs formánsátmenetek révén segítheti az ejteni kívánt magánhangzók azonosítását. Ha azonban a magánhangzók

artikulációs és akusztikai szerkezetében az f0-tól függő változásokat is feltételezünk, amint arra a szerzők második elképzelhető magyarázatként utalnak (és amire gondolni a már ismertett akusztikai elemzések alapján olvasóként is jó okunk lehet), ez a magyarázat nem állja meg a helyét. Ebben az esetben ugyanis a módosult artikulációval ejtett magánhangzók és a mássalhangzók közti átmenetek logikusan csak a módosult képzésű magánhangzót kulcsolhatnák a percepció számára, nem pedig az ejteni kívánt (de valójában nem megvalósított) magánhangzót. Ha tehát az ejteni szándékozott magánhangzók artikulációs-akusztikai megvalósítása módosul az alaphfrekvencia emelésének következtében, a mássalhangzó-környezettől sem várható, hogy az ejteni szándékozott minőségre utaljon, a kontextus legfeljebb a módosult akusztikai szerkezet pontosabb felismerését segíthetné. A szerzők ez utóbbi magyarázatot azonban nem fejtik ki részletesen, így nem tudjuk meg például azt sem, hogy mivel tartják magyarázhatónak azt az elemzésükkor alkalmazott elvet, mely szerint az esetlegesen megváltozott artikulációval ejtett magánhangzókra adandó „helyes válaszként” egyértelműen az íráskép által megjelölt magánhangzót feltételezik. A tanulmány nyomán további kérdésként merül fel az is, hogy mit tapasztalnának a szerzők az általuk vizsgált kondíciókban az igazán magas alaphfrekvenciákon – az ismertetett vizsgálat ugyanis csak egy viszonylag alacsony alaphfrekvenciáig (körülbelül 500 Hz-ig) fedi le a magas hangosztályokba tartozó énekesek által használt frekvenciatartományt (mely a klasszikus értelmezésben 1047 Hz-ig, azaz a „magas céig” terjed).

Scotto di Carlo és Germain (1985) 15 francia magánhangzót elemzett egy percepció vizsgálatban egy koloratúr szoprán izolált ejtésében, az énekes teljes alaphangtartományán rögzítve azokat 220 Hz és 1046 Hz között 17 különböző alaphfrekvencián. A vizsgálatban az /i e ε a α y ø œ u o ɔ/ orális és az /ẽ œ ỹ ã/ nazooralis monoftongusokat használták fel: a szerzők a magánhangzókat izolációban rögzítették és játszották le az észlelési tesztben háromféle randomizált sorrendben. A percepció vizsgálatban négy fonetikailag képzett, de az éneklésben vagy a zenében nem jártas adatközlő vett részt. Az instrukciók szerint az adatközlők feladata az volt, hogy azonosítsák, majd fonetikai szimbólumokkal jelöljék meg az elhangzott magánhangzókat. A teszteket megelőzte egy rövid familiarizációs szakasz is, mely során a kísérletben részt vevő négy adatközlő meghallgatott egy, az énekes által énekelt rövid operarészletet (felvételtől), melyhez a librettót is elolvashatta. Ezt a „tanítási fázist” (mely során a résztvevők

valamelyest megismerkedhettek a stimulusokat szolgáltató énekes hangjával, hangképzésével és énekelt magánhangzóinak sajátosságaival) a rövidtávú memória kondicionálásának tekintve a szerzők úgy vélték, módjuk volt egységesíteni és következetesebbé tenni az egyébként az egyes stimulusokra adott, akár egyénenként is igen változatos megoszlást mutató válaszokat. Vizsgálatukban a szerzők elsősorban az alapfrekvencia hatását vizsgálták, de külön kondícióként a regiszterek hatását is elemezték. E faktor kezelésében Scotto di Carlo és Germain (1985) tanulmánya jól reprezentálja a szakirodalomban egyébként is honoló nagy terminológiai és fogalmi bizonytalanságot. A szerzők ugyanis ötféle regisztert mint alapfrekvencia-tartományt különböztettek meg, rendre alsó (*lower*), alsó-közép (*lower middle*), középső (*middle*), felső-közép (*upper middle*), és felső (*upper*) regisztert. Ezekhez pedig produkciós vagy zöngképzési módokat rendeltek, élesen elkülönítve azokat az előbb említett, az alapfrekvencia-tartományok megnevezését szolgáló fogalmaktól. A kísérletben elkülönített zöngképzési módokat (szintén nem a ma korszerűnek elfogadott terminusokkal élve) a szerzők rendre mellkasi (*chest voice*), alsó közép- (*chest mid voice*), közép- (*mid voice*) és fejhangnak (*head voice*) nevezték<sup>16</sup>. Vélekedésük szerint bár az utóbbiak valóban az előbbiekhöz rendelhetők, ez a hozzárendelés nem egyértelmű, hiszen egyes alaphangtartományok egyes hangjai többféle képzési móddal is képezhetők lehetnek. Épp ebből kifolyólag azonban nehezen értelmezhető, hogy mégis miért, és főként hogy miként használhatták összehasonlításaikban és a

<sup>16</sup> A tanulmányban használt terminológia korszerűtlenségét éppen az mutatja, hogy bár a szerzők állítása szerint a fogalmak fonációs módokat jelölnek, a megnevezéshez mégis az adott módban történő hangadást kísérő szenzoros rezonancia-élmények leírását (ti. pl. *chest voice* = „rezgés a mellkasban”) használják. Mivel ma már a regisztereket egyértelműen gége- és hangszalagszintű, kizárólag a fonációhoz kötődő jelenségeknek tartja a szakirodalom, ezért ezeket a korábban elterjedt megnevezéseket félrevezető mivoltuk miatt egyértelműen kerülni igyekszik. A fonációs módok a kísérletben definiált homályos voltához járul még, hogy az itt *mellkasi hang* és *középhang* terminusokkal megnevezett és egymástól elkülönített produkciós módoknak nem feleltethetők meg egyértelműen a zöngképzés módozataival foglalkozó kurrens szakirodalomban leírt produkciós (azaz zöngképzési vagy rezgés-) módok. A jelenleg legáltalánosabban használt modellek ugyanis (a használati alapfrekvenciák szerint rendre) a) az irreguláris zöngét vagy glottalizációt (*vocal fry*, *creaky voice*, *pulse*, 0. mechanizmus vagy M0), b) a mindennapokban leggyakrabban használt modális zöngét (*modal*, *chest*, *voix mixte [men]*, 1. mechanizmus vagy M1), c) a falzettet (*falsetto*, *head*, *voix mixte [women]*, *mixed*, 3. mechanizmus vagy M3), valamint a d) füttyregisztert (*falsetto*, *flute*, 3. mechanizmus vagy M3) különböztetik meg mint rezgésmódjában eltérő zöngképzési alaptípust (vö. Henrich et al. 2004, Titze 2004, Roubeau et al. 2009, Seikel et al. 2010, Markó 2013). A teljesség kedvéért e kitérőben megjegyezzük azt is, hogy bár a fent felsorolt alaptípusokhoz képest a sokat vizsgált *pressed* vagy „préselt” és *breathy* vagy leheletes zöngképzések hangszinbeli eltéréseket mutatnak, ezek nem az előbbiektől eltérő vibrációs mód révén jönnek létre, így azok a besorolások, melyek a fonációs módot mint hangszalag-vibrációs mechanizmust tekintik, ezeket nem különítik el zöngképzési alaptípusként (vö. pl. Seikel et al. 2010).



statisztikai elemzésekben a produkciós módokat az alapfrekvencia helyett, hiszen arról, hogy a zöngképzést vagy hangszalagrezgést valamely endoszkópián vagy más elven alapuló eszközzel megfigyelték volna, nem számolnak be a szerzők. Feltehető, hogy a fogalmi szétválasztás ellenére a szerzők végül a „produkciós módok” megnevezésen mindvégig az alapfrekvenciákat (és azok tartományait) értik. Mivel így a tanulmányban a „regiszterek” és „fonációs módok” fogalmi folyamatosan keverednek, és ebből kifolyólag a szerzők számos következtetése bizonytalan (vagy azért, mert a zöngképzésből következő, de a frekvenciatartománnyal kapcsolatban értendő, vagy fordítva), az egyes „regiszterekre” vagy „fonációs módokra” kapott eredményeket a jelen értekezésben nem tárgyaljuk. Sorra vesszük most azonban az általuk vizsgált további paraméterekre kapott eredményeket.

Scotto di Carlo és Germain (1985) eredményei szerint a magánhangzóknak az ejtési szándék szerinti azonosítása csökkent az alapfrekvencia emelésével, mely tendenciára mind a magánhangzó-minőség, mind pedig a produkciós mód (vagy talán inkább az alapfrekvencia?), mind pedig a kettő kombinációja befolyással bírt. Mivel azonban sajnos a tanulmányban a legtöbb esetben az egyes „regiszterekre”, azaz alapfrekvencia-tartományokra összevont adatokat közölnek a szerzők, így sok esetben nem tudjuk pontosan meghatározni, mely alapfrekvenciához köthető egy-egy megfigyelt észlelési tendencia megjelenése. A több alapfrekvencia adatait egyetlen adatpontba összesítő elemzésekből és ábrázolásokból azt látjuk, hogy a vizsgált magánhangzóknak az ejtési szándék szerinti azonosítása az  $f_0$  emelésével megközelítőleg fokozatosan csökkent, de olyan módon, hogy egy ponton (a közölt táblázatukból kikövetkeztetve megközelítőleg 349 és 392 Hz, azaz az  $f'$  és a  $g'$  zenei hangok között) az azonosítási százalék drasztikus (62%-ról 38%-ra) esést mutatott, majd a legmagasabb alapfrekvenciákon 9%-ot ért el.

Scotto di Carlo és Germain (1985) az azonosítási tendenciákat a magánhangzók egyes distinktív fonológiai jegyei, mégpedig az ajakműködés (réses és ajakkerekítéses magánhangzók szétválasztása), az állkapocsnyitás (nyílt és zárt magánhangzók szétválasztása), a nazaoorális minőség (nazoorális és orális magánhangzók szétválasztása), valamint a képzéshely, a toldalékcsoában fellépő szűkület helye és a nyelvvállás szerint is elemezték. Ez utóbbi

szempont azonban végül tisztázatlan marad, ugyanis ezekről a jegyekről sem a módszertani bevezetőben, sem pedig másutt nem tudunk meg pontosabbat a tanulmány során.

Az eredmények szerint a fonológiai jegyeknek az ejtési szándéknak megfelelő azonosítási arányai a magánhangzó-minőségek általános azonosítási tendenciáihoz hasonlóan az alapfrekvencia emelésével fokozatosan csökkentek. A vizsgált jegyek között a labialitás tűnt a legkevésbé ellenállónak, melynek az ejtési szándéknak megfelelő azonosítása a legmagasabb alapfrekvenciákon (a 988 Hz-re és az 1047 Hz-re kapott eredmények összevonásával kalkulálva) 1,5%-ig csökkent. Ezt a szerzők a magasabb alapfrekvenciákon általánosan jellemző (és általuk feltehetőleg a felvételek közben megfigyelt) ajakréses ejtéssel magyarázták, melyet az /i/-re és /e/-re magas alapfrekvenciákon talált relatíve magas, az ejtési szándék szerinti azonosítási arányokkal is alátámasztottnak véltek.

Bár nem derül ki egyértelműen, hogy mely magánhangzókat értik a szerzők „zárt” és „nyílt” magánhangzókon (nem használnak ugyanis a köztes fokozatokra vonatkoztatható megnevezéseket, tehát nem egyértelmű az /e ε a ø œ o ɔ/ orális valamint az /ẽ œ ỹ/ nazoorális monoftongusok besorolása), a szerzők egy további eredménye szerint a „zárt” magánhangzók az ejtési szándék szerinti azonosítása az alacsonyabb alapfrekvenciákon magas (46% körüli) volt, a magasabb alapfrekvenciákon az  $f_0$  emelésével egyre csökkent (a legmagasabb alapfrekvencián tapasztalt 4%-ig). Ezzel szemben azonban a „nyílt” magánhangzók az ejtési szándék szerinti azonosítása alacsonyabb alapfrekvenciákon sem volt túl magas (maximum 46%, amit a legalacsonyabb alapfrekvencián tapasztaltak), a magasabb alapfrekvenciákon pedig a zárt magánhangzókhoz hasonlóan a nyíltakra kapott százalékos arány is csökken, ám itt csak 19%-ig. Ezeket az eredményeket a szerzők azzal magyarázzák, hogy amint az korábbi vizsgálatokból tudható, a magas alapfrekvencián ejtett magánhangzók artikulációsan bizonyítottan nagyobb állkapocs-nyitásszöggel valósulnak meg. Ennek megfelelően pedig a szerzők jól beláthatónak tartják, hogy miért nehezebb magas alapfrekvencián a zárt hangzók azonosítása. A szerzők nem térnek ki annak az ellentmondásnak a feloldására, mely e között és az előző következtetésük között feszül: nem értelmezik ugyanis, hogy a magas alapfrekvenciákon alkalmazott nagyobb állkapocsnyitás ellenére miként lehetséges mégis az, hogy az /i/ és /e/ még magas alapfrekvencián is „relatíve jó arányban” volt az ejtés szerint azonosítható (táblázatuk

tanúsága alapján egyébként ez az arány az  $f'$  [698 Hz],  $g'$  [784 Hz],  $a'$  [880 Hz],  $h''$  [988 Hz] és  $c'''$  [1046 Hz] hangokon rendre 50, 25, 16, 25 és 16% az /i/ esetében, valamint 50, 8, 16, 16 és 0% az /e/ esetében). Még jobban megnehezíti az eredmények értelmezését azonban, hogy a szerzők érveikben értelemzavaróan egybemossák a produkciós és percepciós tényezőket. A magasabb alaphangfrekvenciákon ugyanis, a korábbi vizsgálatok szerint, nem a zárt magánhangzók azonosítása nehéz, hanem a hangzók ejtése lehetetlenül el. Hiszen amint azt az előző alfejezetben is láttuk, az énekes magas alaphangfrekvenciákon énekelve és az  $f_0$  további emeléséhez az  $F_1$  értékét az  $f_0$ -ra (vagy afölé) hangolja, amit az állkapocsnyitás szögének növelésének (vagy a nyelvállásfok csökkentésének) segítségével ér el. Ha tehát már eleve nyíltabb hangzókat ejt az énekes, nem a zárt hangzók egyre nehezező „felismeréséről” kell beszélnünk (a szerzők az *identification* szót használják, vö. Scotto di Carlo–Germain 1985: 194), hanem esetleg a zárt magánhangzó-minőség megjelenéséről az észleletben, mert már produkciósan sem (egyértelműen) zárt az adott hangzók minősége.

A nazális jeggyel kapcsolatban (egy eléggé elnagyolt, az összes alaphangfrekvenciát összesítő) összehasonlításban a szerzők azt találják, hogy az ejtési szándék szerint nazoorális magánhangzók nazoorálisként való azonosítása alacsony. Erre magyarázatképp a következő feltételezést fogalmazzák meg: mivel a teljesen nazális vagy nazoorális hangszín az éneklésben nem kívánatos, ezért az énekes igyekezett kerülni azt, mégpedig úgy, hogy lehetőség szerint csak igen kis mértékben, vagy ha ez nem lehetséges, inkább egyáltalán nem nazalizálják a magánhangzókat. (A szerzők szerint ugyanis az énekelve ejtett nazoorális magánhangzók ejtésekor az elsődleges artikulációs cél az orális magánhangzó, melyet az énekesek csak utóbb nazalizálnak – vagy sem, akkor, ha a kismértékű nazalizáció megvalósítását aerodinamikai vagy más okokból nehézségekbe ütközne.)

Végezetül a szerzők a „hibaként” legnagyobb számban megjelenő magánhangzók rangsorát elemzik. Ez alapján megállapítják, hogy a tesztelők az ejteni szándékozott hangzók helyett legnagyobb arányban az /a/ hangzót jelölték meg (mégpedig az összes „hiba”, azaz az ejtési szándéktól eltérő azonosítás 34%-ában). Ezt az eredményt a nyíltság esetében megfogalmazott magyarázatuk ismétlésével értelmezik, azaz felteszik, hogy a nyíltabb artikuláció egyben nyíltabb észleletet is jelent, így a magas alaphangfrekvenciákon produkált énekhang az /a/ ejtésének

kedvez. Mivel azonban ez a lista az összes alapfrekvencián az összes magánhangzóra kapott válaszokat tartalmazza, nem tudjuk meg pontosan, hogy specifikusan mely magánhangzók esetében mely magánhangzóválasz volt jellemző.

Scotto di Carlo és Germain (1985) magyarázatai tehát egyértelműen azon alapszanak, hogy a korábban artikulációs vizsgálatok során demonstrált produkciós tendenciákat lineáris összefüggésbe hozzák a vizsgálatukban talált percepciós adatokkal – így például az artikulációs nyíltság számukra könnyen beláthatóvá teszi a percepciós nyíltság megjelenését. Eredményeik értelmezésekor azonban komoly problémát jelent, hogy hiába közlik a szerzők a „zárt” magánhangzók „nyíltabbként” történt azonosításának tendenciáját, nem tudjuk meg e „zártabb” magánhangzók pontos minőségét. Így nem tudjuk megítélni például azt sem, hogy állításukat olyan magánhangzókra alapozzák-e a szerzők, melyeket (a magánhangzórendszerből következően) egyaránt lehetne nyíltabbként vagy zárttabbként is azonosítani.

Ráadásul az akusztikum és a percepció közti egyszerű lineáris összefüggéseknek a feltételezése sokszor ellentmondásokhoz is vezet a szerzők érvei között, hiszen például – az egyes fonológiai jegyek közötti összefüggések figyelmen kívül hagyásával kezelve a magánhangzókat – az /i/ és /e/ magas alapfrekvencián magasabb arányú az ejtési szándékkal egyező azonosítása értelmezésük szerint egyszer (az ajakműködés szerint) indokoltnak, máskor pedig (a nyíltság szerint) épphogy indoklatlannak tűnik.

Ezen túlmenően fontos megfigyelnünk azt is, hogy a szerzők nem tesznek említést formánshangolásról, sem pedig az akusztikai szerkezetet az alapfrekvencia függvényében érintő más akusztikai változásról, ugyanis az észlelésre kapott eredményeiket közvetlenül és kizárólag a (korábbi leírásokban közölt) artikulációs megfigyelésekkel hozzák összefüggésbe. Ilyen értelemben a vizsgálatukban alkalmazott értelmezési keret talán leginkább a motoros elmélet (vö. *1.1.2.1. alfejezet*) valamifajta változatának lenne tekinthető, de mint ilyen is nagyon leegyszerűsítő és egyúttal sajnos szükségszerűen félrevezető is. Mivel tudjuk, hogy mind az artikuláció és az akusztikum, mind pedig az akusztikum és percepció viszonyát bonyolult összefüggések jellemzik, Scotto di Carlo és Germain (1985) következtetései óvatosan kezelendők.

Bár a fent felsorolt okokból úgy véljük, hogy Scotto di Carlo és Germain (1985) vizsgálatából nem vonhatunk le jól megalapozottnak tekinthető és jól általánosítható konklúziókat, mé-

gis úgy látjuk, hogy a kísérlet számos izgalmas kérdést vet fel, melyeknek ellenőrzése kísérletes vizsgálatot igényel. A szerzők adatai ugyanis rámutatnak például, hogy az alapfrekvencia emelésével nem feltétlenül csak a legnyíltabb magánhangzók észlelete maradhat relatíve magas arányú (lásd az /i/-t és /e/-t), tehát gyümölcsöző lehet a magánhangzó-minőségek minél széleskörűbb összehasonlítása. Ezen túlmenően pedig nem szabad megfeledkeznünk arról sem, hogy ez a vizsgálat tekinthető az első olyan percepciós kísérletnek is, melynek céljai között szerepelt az egyes különböző nyíltsági fokú magánhangzók szétválasztása az észlelési vizsgálatban, ami a korábbi artikulációs-akusztikai eredmények alapján erősen motivált. Így ez a vizsgálat mindenképpen fontos mérföldkő az énekelt magánhangzók észlelését érintő kutatás történetében.

Benolken és Swanson (1990) az angol /i ɪ ε u o ɔ/ monoftongusokat, valamint az /eɪ/ diftongust vizsgálták egy szoprán ejtésében a *hVd* hangkörnyezetben (tehát hat értelmes angol szóban), 13 különböző alapfrekvencián 262 Hz alapfrekvenciától felfelé, két oktáv terjedelemben. Az így rögzített hanganyagot 28 résztvevő segítségével vizsgálták egy percepciós tesztben: az adatközlők egy teremben ülve, hangszórókon keresztül hallgatták az egybefüggő felvételt, és az elhangzott *hVd* hangsorok magánhangzóinak azonosított hangminőségét egy papíron rögzítették. A szerzők azt találták, hogy az alapfrekvencia emelésével egyre csökkent az ejtési szándéknak megfelelő magánhangzó-minőségek azonosításának sikeressége. Ezzel együtt a zárt, első nyíltsági fokú /i/ hangot egyre nagyobb számban azonosították az adatközlők az ejteni szándékozott hangzó minőségénél nyíltabb /ɪ/-ként majd /ε/-ként, valamint a zárt, első nyíltsági fokú /u/ hangot is egyre inkább nyíltabb, azaz /o/ majd /ɔ/ magánhangzókként azonosították a tesztben részt vevők. Ezen felül a szerzők egy másik összefüggésről is beszámolnak. Azt találták ugyanis, hogy a nagyon magas, 1047 Hz alapfrekvenciájú alaphangon a palatális és veláris hangzók csoportjára két egymástól eltérő spektrális és észlelési mintázat volt jellemző. A spektrum tekintetében a velárisok csoportjára az első ( $f_0$ ) és a harmadik ( $3f_0$ ) harmonikus nagyobb, míg a második ( $2f_0$ ) és negyedik ( $4f_0$ ) harmonikus kisebb intenzitása volt jellemző, melyre az észleletben főként veláris hangok, különösen az /ɔ/ és /a/, valamint a palatális /æ/ jelentek meg válaszként. Ezzel szemben a palatálisok esetében az első harmonikustól a negyedikig fokozatosan csökkenő amplitúdó volt jellemző, amire az észleletben (a

válaszok között) veláris és palatális hangzók egyaránt megjelentek – a velárisoknál tapasztalt szórásnál nagyobb szóródást mutatva.

A szerzők értelmezése szerint a beszédhangok az ejtési szándéknál nyitottabb hangokként való azonosítása a korábban leírt  $F1 : f_0$  hangolási tendenciákkal egybevágó eredmény, vélekedésük szerint tehát az akusztikum és a percepció összefüggése a magas alapfrekvenciájú énekelt magánhangzók esetében lineáris, ahol a nyíltabb akusztikai vetület nyíltabb percepciós vetületet is implicál. A magas alapfrekvenciákon talált kétféle spektrális mintázatból ezen túlmenően pedig arra is következtettek, hogy az 1047 Hz alapfrekvencián 500 Hz alatt megtalálható négy harmonikus ( $f_0$ ,  $2f_0$ ,  $3f_0$ ,  $4f_0$ ) elegendő információt hordoz(hat) ahhoz, hogy a palatális és veláris magánhangzócsoporthoz még ilyen magas alaphangon is (bizonyos mértékig) elkülöníthetőek maradjanak. Sajnos azonban a szerzők adósok maradtak azzal a nem magától értetődő magyarázattal, hogy milyen módon volna párhuzamba hozható a magas alapfrekvencián talált spektrális mintázat az alacsonyabb frekvenciákon jellemző formánsszerkezettel. Nem adnak tehát választ (vagy javaslatot) arra a kérdésre, hogy vajon miként feleltetheti meg a beszédpercepció a magasabb alapfrekvenciákon megvalósult énekelt magánhangzók (alulmintavételezett és a formánshangolás miatt megváltozott) spektrális szerkezetét az alacsonyabb alapfrekvenciákon tipikusan megszokott spektrális szerkezeteknek.

Hollien és munkatársai (2000) az egymástól artikulációs és akusztikai tekintetben legmesszebb elhelyezkedő (amerikai angol) /i u a/ magánhangzók észlelését vizsgálták 13 haladó nő és 5 férfi énekes ejtésében „magas” és „mély” alaphangokon, „halk” és „hangos” hangadásban. A nők esetében az „alacsony” alaphangok a *G* (98 Hz) vagy az *A* (110 Hz) zenei hangokat jelentették, a „magas” alaphang esetében pedig mindössze annyi megszorítást tettek a kutatók, hogy az énekeseket legalább egy *a*'' (880 Hz), de lehetőség szerint magasabb hang éneklésére kérték (az énekesek tehát a megjelölt hangokból tetszőlegesen választhattak, így az alapfrekvencia az egyes énekesek ejtésében a „mély” és „magas” kondíciókon belül nem feltétlenül egyezett). Az eredmények szerint a „mély” kondícióban a női énekesek által elért alapfrekvencia átlagosan 200 Hz körül, a „magas” kondícióban mérhető átlagos alapfrekvencia pedig 900 és 1000 Hz között mozgott. A percepciós teszthez a kutatók a magánhangzók közepéből (a magánhangzók tiszta fázisának tekinthető) 1,5 másodperces szeleteket vágtak ki, majd a hang-

zókat randomizált sorrendben játszották le a 38 képzett (beszédtanár, fonetikus végzettségű, illetve beszédterapeuta-hallgató) és 12 nem képzett résztvevőnek. A percepció vizsgálatot akusztikai elemzés kísérte, a szerzők az énekelt magánhangzók első két formánsának frekvenciaértékét mérték (hagyományos, spektrális elemzés segítségével).

A percepció teszt eredményei szerint bár a hangerősség bizonyos mértékben másként hatott az egyes hangokra (az /i/ és /a/ nagyobb hangerő mellett inkább az ejtési szándék szerint volt felismerhető, míg az /u/-ra a hangerősség éppen fordított hatást gyakorolt), ez a különbség (statisztikailag) nem volt számottevő. Az ejtési szándék szerinti azonosítás (a hangerősség figyelembevétele nélkül átlagolva) az /i/ esetében az alacsonyabb alapfrekvencián 26,7%, majd a magasabbon 2,2% volt, az /u/ esetében 38%, majd 1,8%, az /a/ esetében pedig 74,2, majd 45,5%. Ezek alapján a gyakorlatilag minden alapfrekvencián az ejtési szándéktól nagy arányban eltérő azonosítást mutató eredmények alapján (melyekben az  $f_0$  emelésével csökkenés is látható volt) a szerzők azt a következtetést vonták le, hogy adataik a korábbi eredményekkel egybevágnak. Mivel azonban az akusztikai adatok alapján az alacsonyabb alapfrekvenciákon még nem lépett fel  $F_1 : f_0$  hangolás, az itt talált alacsony azonosítási arányokat nem tudják magyarázni. A magas alapfrekvenciákon talált, az ejtési szándéktól eltérő azonosítási tendenciákról kiemelik, hogy azokat a beszédben magasabb  $F_1$ -gyel rendelkező /a/ válaszok dominálják, ugyanakkor táblázataikból jól látható, hogy a fonológiai laza /u/ válaszok is nagyarányúak (15-16% körüliek), ami valamivel a véletlen szintje fölött van (az ugyanis a nyolc megadott lehetséges válasz esetén 12,5%) a magas alapfrekvenciákon mind a három ejteni szándékozott magánhangzó-minőség esetében. Míg az /a/ válaszokat a korábbi tanulmányokban megszokott módon a szerzők a nyíltabbként ejtésből és a magasabb  $F_1$ -értékből következő nyíltabbként azonosítással, azaz az akusztikum és a percepció egyszerű lineáris összefüggéséből magyarázzák, ez utóbbi eredmény értelmezésére nem tesznek kísérletet. Mivel az akusztikai elemzésekben a kontroll feltételt, azaz az összehasonlítás alapját nem a beszéd, hanem az „alacsony” kondíció adta, nem tudjuk meg, hogy ezen az alapfrekvencián voltak-e eltérések a beszélők beszélt és énekelt magánhangzói között. Annyi azonban bizonyos, hogy az alacsonyabb alapfrekvenciákon mért  $F_1$ - és  $F_2$ -értékekhez képest a „magas” kondícióban jóval magasabb  $F_1$ - és  $F_2$ -értékeket mértek a szerzők (természetesen minden magánhangzó esetében az

$f_0$  és a  $2f_0$  frekvenciaértékén, hiszen a hagyományos spektrális elemzésekkel mindösszesen az erősített felharmonikusok frekvenciaértéke állapítható meg, nem pedig a formánsfrekvencia valós középértéke).

Az eddig ismertetett munkákhoz hasonlóan ez a tanulmány is hordoz némi ellentmondást az eredmények értelmezésében. Bár Hollien és munkatársai (2000) látszólag úgy vélik, hogy a magánhangzók az ejtési szándéktól eltérő minőségű magánhangzóként történt azonosítása az akusztikumnak a magas alapfrekvencia és az  $F_1$  hangolása<sup>17</sup> miatt megváltozott jellegéből ered, mégis, a tanulmány több pontján utalnak arra, hogy a tapasztalt alacsony azonosítási tendenciákon (a koartikulációs hatások miatt) minden bizonnyal „javítana” a mássalhangzó-környezet jelenléte. Ennélfogva nem teljesen világos, hogy a szerzők feltételezése szerint a mássalhangzó-környezet a magánhangzók ejtését vagy inkább csak azonosítását befolyásolja-e.

Az énekelt magánhangzók azonosításának vizsgálatával kapcsolatosan összegzőképpen a következőket mondhatjuk el. A vizsgálatok eredményeiben közös, hogy az ejteni szándékozott magánhangzók azonosításának az alapfrekvencia emelésével csökkenő tendenciáját mutatták ki. Azon vizsgálatok, melyek a magánhangzót körülvevő mássalhangzó-környezet pozitív hatását kívánták bebizonyítani, azaz azt feltételezték, hogy az alapfrekvencia emelése következtében előálló artikulációs/akusztikai változások a formánsátmenetekben kódolt információ miatt perцепción(?) kiküszöbölhetők, várakozásaiknak megfelelő eredményeket találtak. Magyarázataikat alaposabban szemügyre véve azonban azt találtuk, hogy azok ellentmondásosak, és több ponton vegyítik a magánhangzók produkciós és perцепciós vetületeivel kapcsolatos érveket és feltételezéseket. E tanulmányok szerzői ugyanis nem foglalnak egyértelműen állást abban a kérdésben, hogy az  $F_1$  a megemelt alapfrekvenciára hangolását elfogadják-e a magas alapfrekvencián énekelt hangzók esetében, és ha igen, miként értelmezendő a mássalhangzók az általuk javasolt „kompenzációs” hatása az ily módon (az alacsony alapfrekvenciákhoz képest) megváltozott akusztikai szerkezetű magánhangzók észlelésében.

---

<sup>17</sup> A szerzők valójában nem említik az  $F_1 : f_0$  hangolást (feltehetően azért, mert a „formáns” fogalmát az akusztikai kimenetben mérhető energiamaximumra értik), így mindösszesen csak azzal kapcsolatos spekulációikat fogalmazzák meg, hogy a megemelt  $f_0$  miatt az  $F_1$  törvényszerűen magasabb lesz, mégpedig, szintén törvényszerűen az  $f_0$  értékével egyező.



Bár az idézett tanulmányokban a legtöbb esetben a percepció adatokhoz nem kapcsolódtak akusztikai elemzések, azokban az esetekben, ahol a percepció adatokat valamilyen jellegű akusztikai adatokkal is összevetették a szerzők, az F1 az f0-ra hangolásának jeleit (vagy az f0 és F1 „összeolvadását”) vélték felfedezni magasabb alaphangfrekvenciákon (vö. Hollien et al. 2000), valamint a veláris és palatális magánhangzócsoporthoz jellemző, eltérő akusztikai és észlelési mintázatokat (vö. Benolken–Swanson 1990).

Következtetéseik között az énekelt magánhangzók percepcióját vizsgáló szerzők gyakorlatilag minden esetben valamely az artikuláció és a percepció, illetve az akusztikum és a percepció között fennálló egyszerű, direkt és lineáris összefüggések feltételezéséből indultak ki, melyek értelmében az F1 : f0 hangolás következtében nagyobb állkapocsnyitással, illetve magasabb F1-gyel ejtett magánhangzók észlelésében is az „eredeti” magánhangzó-minőségeknél nyíltabb magánhangzók felismerését tartották feltételezhetőnek. Az ezekkel az elvárásokkal egybevágó eredményeiket tehát indokoltnak (és további magyarázatot nem igénylőnek) találták. Mindebből tehát úgy tűnik, hogy a szakirodalom javaslata szerint a magas alaphangfrekvenciákon énekelt magánhangzók azonosítása, az azonosításra jellemző tendenciák triviálisak. Mégis okunk van feltételezni, hogy ezek a javaslatok túlságosan leegyszerűsítőek, és az esetek egy részében legalábbis biztosan nem elegendőek a percepció működésének értelmezéséhez.

Egyrészt, ahogyan azt az iménti ismertetések során is láttuk, az egyes vizsgálatokban találni az ezzel az egyszerű lineáris összefüggéssel nem jól magyarázható tendenciákat is (vö. pl. az /i/ és /e/ az ejtési szándéknak megfelelő azonosítása magas alaphangfrekvenciákon Scotto di Carlo és Germain 1985 kutatásában). Másrészt pedig nem szabad megfedkeznünk arról sem, hogy a magánhangzók és magánhangzószzerű hangok feldolgozásának egyes kísérletes eredményei szerint a legtöbb percepció helyzetben nem elegendő egyetlen formáns frekvenciaértékéből kiindulnunk a beszédhangok észleleti vetületének megértéséhez. Ezen vizsgálatok értelmében ugyanis például a magánhangzók nyíltságának észlelete (mely az énekelt magánhangzók esetében az F1 frekvenciaértékének intenciózus módosításai miatt a legégetőbb kérdésnek látszik) nem pusztán az első formáns frekvenciaértékének, hanem inkább a spektrum prominenciái közötti relációknak a függvénye (vö. a jelen értekezés 1.1.2.2. és a 2.5. *alfejezetét*).

Mindent egybevetve úgy látjuk, hogy az énekelt magánhangzók percepciója az artikulációs és akusztikai megvalósulás vizsgálatához képest alulreprezentált, méltatlanul mellőzött kérdés a szakirodalomban. Ennek a hiányosságnak az egyik aspektusa, hogy az énekelt magánhangzók észleléséről eleve csak viszonylag kisszámú kísérleti eredményt találunk az irodalomban. Egy másik aspektus, hogy a vizsgálatok java az artikuláció és észlelés vagy az akusztikum és észlelés között egyszerű, direkt és lineáris összefüggéseket feltételezett, ily módon tehát ezek a vizsgálatok feltehetően túlzó módon leegyszerűsítették az énekelt magánhangzók azonosításának kérdését – miközben láttuk azt is, hogy ezek a leegyszerűsítő összefüggések több ponton nem magyarázzák a talált tendenciákat. Egy harmadik aspektus, hogy a kísérletek legtöbbje a vizsgált kérdés megválaszolásában módszertanilag több, az eredmények értelmezését is megnehezítő problémát tartalmazott. Egy negyedik aspektus, hogy a vizsgálatok nem igyekeztek az egyes nyelvekre jellemző magánhangzókészlettel szorosabb összefüggésbe hozni az énekelt magánhangzókra jellemző percepciós tendenciákat, így valójában igen keveset tudtunk meg a percepciós működésekről a magas alapfrekvenciájú énekelt magánhangzók azonosításában. A nem az ejtési szándék szerinti észlelési tendenciákat ugyanis a legtöbb vizsgálatban nem (vagy nem megbízhatóan) elemezték a szerzők. Egy ötödik aspektus, hogy bár a magánhangzók a hangkörnyezetben ejtett megvalósításainak azonosítását és ilyen értelemben a hangkörnyezet hatását a szakirodalom megválaszolt kérdésnek látszik kezelni, úgy véljük, a problémakörben adott válaszok korántsem megnyugtatók, és további kérdéseket, alaposabb vizsgálatot érdemelnek. Végezetül pedig a hatodik szempont, hogy amint láttuk, igen ritka az olyan vizsgálat, melyben az akusztikai és percepciós eredmények összeolvasásán keresztül képet kaphatnánk arról, hogy pontosan milyen hatások és hogyan is befolyásolják a magas alapfrekvenciákon énekelt magánhangzók azonosítását. Noha tudvalévő, hogy a magas alapfrekvenciájú magánhangzók spektrális alul-mintavételezettségének köszönhetően mind a rezonanciaadatok, mind pedig az akusztikum percepciós vetülete a beszéd és beszédészlelés vizsgálatában megszokottnál bonyolultabb kérdés, mégis úgy véljük, hasznos volna egy szisztematikus összehasonlító vizsgálat, mely legalábbis további feltételezések megfogalmazását lehetővé tenné.

A fentiek alapján úgy látjuk, hogy máig nem tartható megválaszoltnak az a kérdés, hogy pontosan hogyan is működik az emberi beszédpercepció egy olyan extrém percepciós helyzet-

ben, mint a magas alapfrekvenciájú énekelt magánhangzók azonosítása. Kísérleteinket így dominánsan ennek a kérdésnek szenteltük a magyar magánhangzók esetében a magánhangzók akusztikai szerkezetének vizsgálatával kiegészítve.

### **1.3. A jelen kutatás célja, kérdések, hipotézisek**

A jelen értekezésben bemutatásra kerülő vizsgálatok célja a magas alapfrekvenciájú éneklésben megvalósuló magyar magánhangzók akusztikai és percepciójás jellemzőinek leírása. A kísérleteink egy részében az énekelt magánhangzók produkcióját érintő kérdésekre kerestük a választ, a további kérdéseket pedig a magánhangzók percepciójával kapcsolatban fogalmaztuk meg. Kérdéseinket az előzőekben bemutatott kísérletek nyomán felmerült problémákra és felvetésekre alapoztuk, és négy kérdés csoportba rendeztük.

Az énekelt magánhangzók akusztikumát érintő kérdések:

1. Hogyan írhatók le a magyar magánhangzók érintő produkciós változások az alapfrekvencia függvényében, különösen akkor, ha az  $f_0$  magasabb, mint az adott magánhangzónak a beszédben megjelenő első formánsa? Mi jellemzi a spektrális maximumokat, és mi jellemzi a toldalékcsoportban az artikuláció során megjelenő ( $F_1$  és  $F_2$ ) üregi rezonanciákat? Mivel a kérdések megválaszolására a hagyományos, Fourier-transzformáción alapuló spektrális elemzésen túl módszertani újítások alkalmazása is szükséges, a kérdéshez kapcsolódó további megoldandó problémát jelent egy megfelelő módszertani eljárás kidolgozása és tesztelése is.

Az énekelt magánhangzók percepcióját érintő kérdések:

2. Mi jellemző az énekelt magyar magánhangzók észlelési tendenciáira? Hogyan változik az ejtési szándék szerinti azonosítás az alapfrekvencia emelésével? Igazolható-e, hogy a magas alapfrekvenciákon csak a nyíltabb ejtésű magánhangzók az ejtésnek megfelelő azonosítása jellemző? Mi jellemzi az ejtési szándéktól eltérő azonosításokat? Igazolható-

e a produkció és a percepció közti lineáris összefüggés, azaz az alapfrekvencia emelésével megjelenő egyre nyíltabb magánhangzó-minőségek azonosítása?

3. Miként hat a mássalhangzó-környezet az énekelt magánhangzók azonosítására? Milyen hatása van a koartikulációs formánsátmenetekben rejlő akusztikai információnak a magánhangzók észlelésére? Miként hat (és a formánsátmeneteknél tapasztalt hatástól elkülöníthető-e) a mássalhangzó-környezet akkor, ha a mássalhangzók és a magánhangzó kapcsolatából előálló szekvenciák értelmes szavakat alkotnak (tehát az észlelést a felsőbb nyelvi szintekhez kapcsolódó feldolgozási folyamatok aktivációja is segíti)?
4. Miként hat a zöngéindítás az énekelt magánhangzók elkülönítésére/azonosítására? Alkalmazható-e az énekelt magánhangzók esetében az a hangszerek hangszínének megkülönböztetésében talált eredmény, mely szerint a hangindítás mint akusztikai kulcs jelenléte növeli a hangszín felismerésében nyújtott percepciós teljesítményt?
5. Különbözik-e bármilyen módon a naiv, énektechnikailag képzetlen hallgatók és az énekesek percepciója – sikeresebbek-e az énekesek az egyes énekelt hangzóknak az ejtési szándék szerinti azonosításában? A kérdés vizsgálata azt célozza, hogy a lehetőségek szerint legalaposabban körüljárjuk a produkciós változások percepciós kompenzációjának lehetséges módozatait.

A vizsgálandó kérdéseinkkel kapcsolatban három hipotézist fogalmaztunk meg:

**H1:** Az énekesek a magas alapfrekvenciákon énekelt magyar magánhangzók első formánsát a megemelt  $f_0$  értékére (vagy valamivel afölé) hangolják akkor, ha egyébként az  $f_0$  értéke magasabb lenne, mint a magánhangzóra a beszédben jellemző első formáns értéke. Ennek értelmében az  $F_1 : f_0$  megjelenése az éneklésben a magánhangzók nyíltsági fokának (illetve beszédben jellemző  $F_1$  értékének) függvénye.

**H2:** Mivel magas alapfrekvenciákon az énekelt magánhangzóknak a frekvenciaszerkezete és a spektrális felbontása is megváltozik, a magas alapfrekvencián énekelt magyar magánhangzóknak az ejtési szándék szerinti azonosítása az alapfrekvencia emelésével törvényszerűen csök-

ken. Az  $f_0$  és  $F_1$  csökkenő távolsága miatt a magas alapfrekvenciákon ejtett magánhangzók akusztikai megjelenése az ejtési szándéktól eltérő azonosításokban inkább a zárt hangzók felismerését facilitálja.

**H3:** Mivel magas alapfrekvenciákon az énekelt magánhangzók frekvenciaszerkezete megváltozik, továbbá a magánhangzók spektrális felbontása (a környező mássalhangzókkal alkotott szonoráns hangátmenetek spektrális felbontásával együtt) csökken, a mássalhangzó-környezet a formánsátmenetekben rejlő akusztikai kulcsok révén nem fejt ki pozitív hatást a magánhangzók az ejtési szándék szerinti azonosítására. Ugyanez érvényes az izolált ejtésű magánhangzók zöngéindítására is. A mássalhangzó-környezet pusztán abban az esetben facilitálhatja az ejtési szándék szerinti azonosítást, ha az azonosítandó magánhangzót tartalmazó hordozó hangsorok olyan értelmes szavak, melyek száma kisebb, mint a válaszként szóba jöhető összes magyar magánhangzó száma.

## **2. Kísérletek**

### **2.1. A magyar magánhangzók ejtése és észlelése az éneklésben**

Az első vizsgálatunkban a magyar magánhangzók akusztikai és perцепciós vetületét vizsgáltuk meg a szopránéneklésben az alapfrekvencia függvényében. Első célunk az volt, hogy feltárjuk a magas alapfrekvencián énekelt magánhangzóknek a hagyományos (Fourier transzformáción alapuló) akusztikai elemzés segítségével kinyerhető jellemzőit, illetve hogy megállapítsuk az alapfrekvencia emelésével megfigyelhető változásokat. A második célunk az ejteni szándékozott magánhangzók észlelésére az alapfrekvencia emelésével megjelenő tendenciák vizsgálata volt.

Amint azt már az előző fejezetben több pontot is kiemeltük, a formánsok hagyományos (Fourier-transzformáción alapuló) detekciója a magas alapfrekvenciák esetében problematikus: a magas alapfrekvencián ejtett magánhangzók spektrális alul-mintavételezettsége miatt ugyanis az akusztikai kimenetben nem egyértelműen az ejtésekor jellemző üregi sajátrezonanciák középértéke határozható meg, hanem inkább az egyes rezonanciák sáv szélességébe tartozó felharmonikusok frekvenciája. Szükségesnek tartjuk azonban annak alapos felmérését, hogy milyen lehetőségei vannak a hagyományos elemzésnek, valamint szükséges, hogy megállapítsuk az akusztikai kimenet maximumhelyeit az észlelés szempontjából is – az emberi beszédperceptió számára ugyanis csak ezek az információk állnak rendelkezésre az egyes hangzómínőségek detektálásához.

#### **2.1.1. Anyag, módszer és kísérleti személyek**

A vizsgálatunk anyagát a hosszú vagy fonémakategória-váltás nélkül hosszan ejthető magyar /ɒ a: ɛ e: i: o: ø: u: y:/ magánhangzók képezték. Mivel a rövid magánhangzók ejtése az éneklésben problematikus lehet – hiszen felmerülhet, hogy az énekes a hosszú magánhangzókkal való kontrasztjuk miatt a rövideket igyekszik túlságosan erősen redukált időtartamban

megvalósítani, így esetleg nem a hangképzési kívánalmaknak megfelelő hangszínezettel ejteni azokat –, ezért vizsgálatukat itt is, és másutt is kihagytuk a kísérleteinkből. Ezzel együtt mégis biztosított, hogy a jelen kísérletben az összes magyar magánhangzó-minőség vizsgálatára módunk volt, ugyanis az ebből a felsorolásból kimaradt nyelvileg rövid magánhangzókat korábbi eredmények és a szakirodalomban hagyományos vélekedés alapján mint hosszú párjuktól csak időtartamukban eltérő hangzóminőségeknek tekinthetjük (vö. 1.1.1.4. *alfejezet*).

A mássalhangzó-környezet bizonyítottan megkönnyíti a magánhangzók felismerését (vö. 1.1.2.2. *alfejezet*), továbbá egyes énekesek beszámolóiból és énektechnikai szakkönyvek alapján is tudható, hogy a magánhangzót övező nazális mássalhangzók a megfelelő rezonanciaérzet biztosítása miatt megkönnyítik az énekes számára a magas alapfrekvenciájú hangadást (vö. Kerényi 1959, Váradi 2010). Éppen ezért a jelen kísérletben a magánhangzókat nazális hangkörnyezetben, az *mVn* hangkapcsolatban ejtve (értelmes vagy értelmetlen), egyszótagú szavakban (logatomokban) elemeztük (pl.: *man*).

A teszthangsorokat egy professzionális szoprán énekesnő (életkora 50 év) énekprodukciónak rögzítettük (csendesített szobában, Zoom H4 típusú irányított mikrofonnal, 44,1 kHz-en, 16 biten) számára kényelmes hangerősség mellett, olyan módon, hogy az énekes a hangsorokat kitartva (változatlan alapfrekvencia mellett), közel 1,5-2 s időtartamban ejtette ki beszélve (átlagosan 200 Hz alapfrekvencián), valamint énekelve az 500, 550 és 650 Hz alapfrekvenciákon. Az alapfrekvenciák eléréséhez minden felveendő blokk előtt referenciahangot juttattunk az énekesnő fülébe fülhallgatón keresztül. Az alapfrekvencia emelésének mértékét annak megfelelően választottuk, hogy a kísérletben részt vevő énekesnő kényelmesen, megerőltetés nélkül tudja végigénekelni a hangsorokat az összes alapfrekvencián.

A percepciók tesztben a szoprán énekessel rögzített hangsorok magánhangzóinak azonosítását teszteltük tíz nem képzett (tehát nem a beszéd- vagy énekhanggal foglalkozó) adatközlő részvételével (4 férfi és 6 nő, 21 és 25 év közöttiek, átlagéletkoruk 22 év). A tesztelők feladata az volt, hogy azonosítsák az elhangzó hangsorokban felismert magánhangzót. Ehhez egy válaszlapot biztosítottunk, melyen feltüntettük az elhangzó hangsorok mássalhangzóvázát, így a válaszadóknak csak a köztes magánhangzó minőségét kellett megjelölniük helyesírásban. A kísérlet előtt minden adatközlőnek felhívtuk a figyelmét, hogy a kísérletben értelmes és értel-

metlen szavak egyaránt előfordulhatnak, így ne próbáljanak értelmes szavakat felismerni az elhangzó szótagokban, csak a magánhangzó minőségére figyeljenek. A percepció tesztben a (9 magánhangzó  $\times$  4 alaphfrekvencia =) 36 teszthangsorot mind magánhangzó-minőség, mind pedig alaphfrekvencia szerint randomizált sorrendben prezentáltuk a résztvevőknek disztraktor hangsorokkal vegyesen (a disztraktorok különböző mássalhangzó-környezetben beszélve vagy énekelve ejtett magánhangzók voltak) fülhallgatón keresztül.

A felvett hangsorok magánhangzóiban ez után megállapítottuk az énekelt magánhangzók első két formánsának feltételezhető értékét, azaz detektáltuk az akusztikai kimenetben megfigyelhető maximumok frekvenciahelyeit a Praat programban (Boersma–Weenink 2009). Az adatokon statisztikai próbákat végeztünk az SPSS 13.0 program segítségével: Shapiro–Wilk-próbával ellenőriztük a normális eloszlás kritériumát, parametrikus adatok esetén egyváltozós varianciaanalízist alkalmaztunk (ANOVA), nem parametrikus adatok esetén pedig  $\chi^2$ -próbát vagy Mann–Whitney U-próbát.

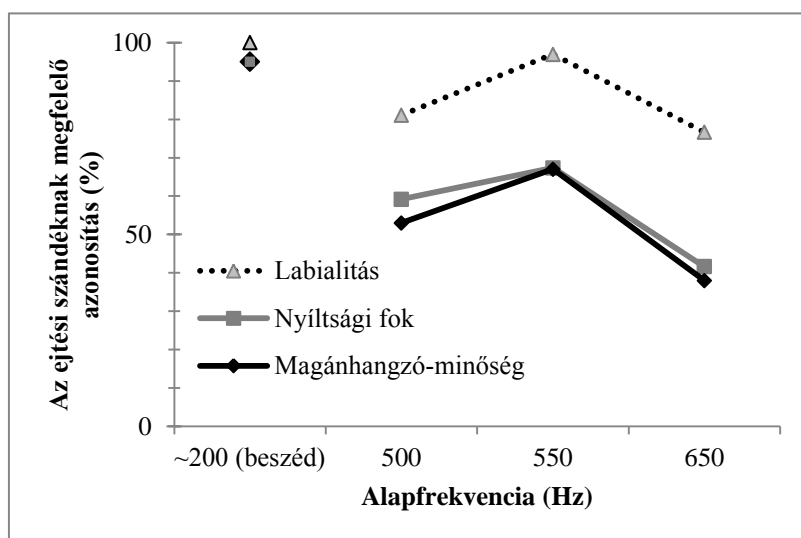
### **2.1.2. Eredmények**

#### **2.1.2.1. Észlelési vizsgálat**

A percepció teszt átlageredményeit a 10. ábrán foglaltuk össze. Az ábra tanúsága szerint az egyes magánhangzók az ejtési szándék szerinti azonosítása – az elvárásoknak megfelelően – csökkent az alaphfrekvencia emelésével, tehát minél magasabb alaphfrekvencián énekelte az egyes magánhangzókat az énekes, annál kevésbé ismerték fel azokat a hallgatók az ejtési szándék (azaz az énekesnek a produkció felvételkor mutatott stimuluslista) szerint. Az ábra az egyes alaphfrekvenciákon ejtett magánhangzók, valamint a magánhangzók képzési jegyeinek összesített azonosítási százalékait mutatja: egy adatpont az egyes alaphfrekvenciákon ejtett összes magánhangzóra adott összes válasza utal, azaz az egyes pontok a kilenc magánhangzóra a tíz adatközlő által szolgáltatott válaszokat összesítik. Az első, elkülönítve megjelenített adatpont mutatja a beszéd módra kapott azonosítási adatokat. Az elkülönítés oka, hogy ezek az adatpontok nem csak alaphfrekvenciájában, hanem produkció módjában is eltérő stimulusokra kapott eredményeket mutatnak.



Az ábráról leolvasható, hogy míg beszédben a magánhangzók az ejtési szándéknak megfelelő azonosítási aránya közel 100%, a kísérletben legmagasabb, 650 Hz alapfrekvencián ejtett magánhangzókat már csak 38%-ban azonosították az adatközlők az ejtési szándék szerint. Az ábra azonban egyúttal azt is mutatja, hogy az azonosítási arányok csökkenése nem teljesen fokozatos: 500 Hz alapfrekvencián a beszédhez képest csökken az ejtési szándék szerinti azonosítás aránya, 550 Hz alapfrekvencián azonban ismét megnövekszik, míg 650 Hz alapfrekvencián újból csökkenést látunk.



10. ábra: Az egyes magánhangzó-minőségek az ejtési szándék szerinti azonosítása, valamint az ejteni szándékozott nyíltsági fok és ajakműködés az ejtési szándék szerinti azonosítása az alapfrekvencia függvényében

Az eredmények szerint a labialitás jegy a legellenállóbb az alapfrekvencia emelésére. Ennek értelmezéséhez természetesen figyelembe kell vennünk, hogy a kilenc vizsgált magyar magánhangzó közül négy ajakréses, míg öt ajakkerekítéses, a vizsgált anyag az ajakműködés szerint tehát kiegyenlítettnek tekinthető. A labialitás és a magánhangzók nyíltsági fokának összevetéséből kiderül, hogy a nyíltság a labialitásnál szignifikánsan kisebb mértékben rezisztens a hangmagasság emelésére (ANOVA,  $F(2) = 8,34$ ;  $p = 0,02$ ), ami azt jelenti, hogy az alapfrekvencia emelésével az egyes magánhangzókat egyre kisebb mértékben tudták az adatközlők az

ejtési szándék szerinti nyíltsági fokú magánhangzóként azonosítani (szemben a beszéddel, ahol ez az arány közel 100%).

Az egyes magánhangzók az ejtési szándék szerinti azonosításának arányát az alapfrekvencia függvényében a 2. táblázat foglalja össze. A korábbi, az észlelést az artikulációval és percepcióval egyenes összefüggésben feltételező (az előző fejezetben bemutatott) szakirodalom szerint az egyre magasabb alapfrekvenciák a nagyobb nyíltsági fokú magánhangzók azonosításának kedveznek – fontos felidézni azonban azt is, hogy ennek az elvárásnak egyáltalán nem minden empirikus adat felelt meg. A korábbi vizsgálatokban ezzel kapcsolatban azt jegyezték fel, hogy a magasabb alapfrekvenciákon nagy számban jelentek meg az /i:/ válaszok, az ejtési szándéknak megfelelő vagy attól eltérő válaszként egyaránt. A 2. táblázat szerint a jelen vizsgálatban a legmagasabb alapfrekvencián nem a legnyíltabb /ɒ/ és /a:/ ejtési szándék szerinti azonosítása volt a legmagasabb, hanem az első nyíltsági fokú /i:/ és /y:/, valamint a harmadik nyíltsági fokú /ɛ/ hangoké. A negyedik nyíltsági fokú /ɒ/ esetében az ejtési szándék szerinti azonosítás fokozatos csökkenést mutatott, az /a:/ esetében pedig az átlagadatokban is látható ugrással megszakított, de egyébként csökkenő tendenciát.

2. táblázat: Az egyes ejteni szándékozott magánhangzók az ejtési szándék szerinti azonosítása az alapfrekvencia függvényében

Alapfrekvencia (Hz)	Az egyes ejteni szándékozott magánhangzók az ejtési szándék szerinti azonosításának aránya (%)								
	/a:/	/ɒ/	/o:/	/u:/	/ø:/	/y:/	/ɛ/	/e:/	/i:/
Beszéd (~200)	100	90	100	95	100	100	100	100	71
500	22	77	62	85	4	50	73	24	90
550	100	71	55	43	50	71	95	60	58
650	26	9	30	43	0	65	67	38	67

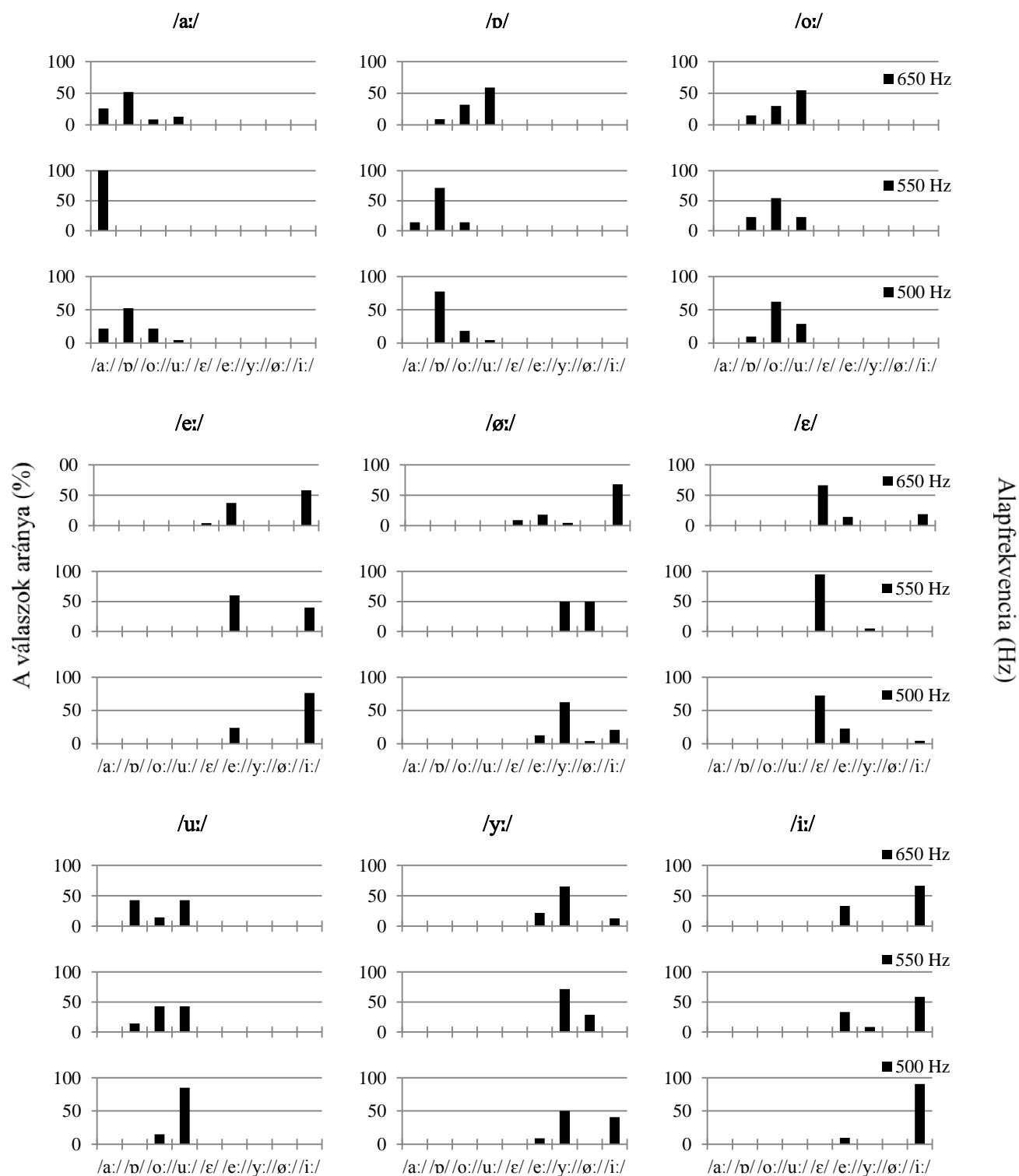
Az észlelési tendenciákat jellemző, az ejtési szándéktól eltérő azonosítások elemzéséhez előbb konfúziós mátrixokba gyűjtöttük az összes alapfrekvencián kapott összes magánhangzóválaszt (vö. 3. táblázat), majd ezeket az összesített adatokat – a könnyebb olvashatóság kedvéért – grafikusán is ábrázoltuk (11. ábra).

A példaként közölt konfúziós mátrix a következőképpen olvasandó: az ejteni szándékozott /a:/ hangzót az adatközlők a válaszok 26,1%-ában /a:/-ként, 52,2%-ban /ɒ/-ként, 8,7%-ban /o:/-ként, 13%-ban pedig /u:/-ként azonosították (a mátrix átlójáról tehát az ejtési szándék szerinti azonosítások olvashatók le, míg az azon kívül eső elemek a nem az ejtési szándék szerint azonosított magánhangzóválaszokat mutatják). A grafikon a számadatoknak megfelelően a válaszok százalékos megoszlását ábrázolja a következőképpen: az /a:/ magánhangzóhoz tartozó oszlop első sorában elhelyezkedő panelen (mely a legmagasabb alaphangfrekvencián kapott válaszokat mutatja) az oszlopok magassága ( $y$  érték) az  $x$  tengelyen feltüntetett magánhangzók mint válaszok megjelenésének százalékos aránya.

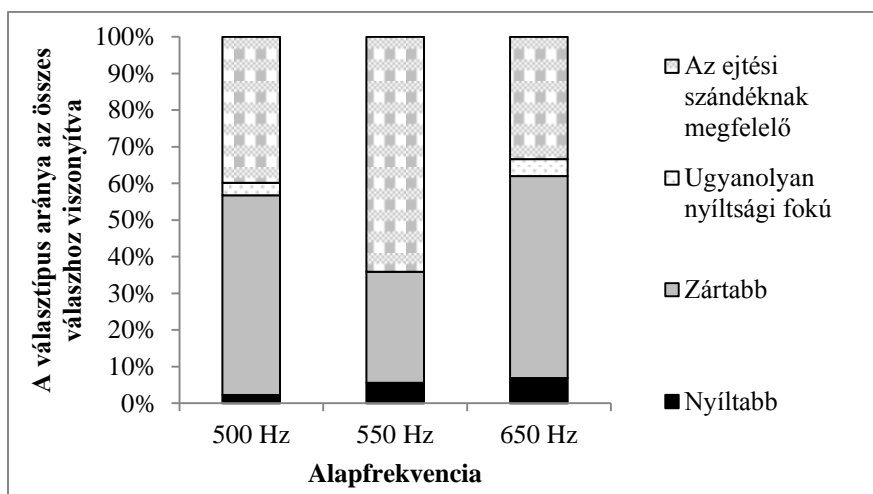
3. táblázat: A legmagasabb alaphangfrekvencián (650 Hz) ejtett magánhangzókra adott válaszok megoszlása a lehetséges válaszok között

		Válasz (%)								
		/a:/	/ɒ/	/o:/	/u:/	/ø:/	/y:/	/ɛ/	/e:/	/i:/
Az ejteni szándékozott magánhangzó	/a:/	26,09	52,17	11,76	13,04	0	0	0	0	0
	/ɒ/	0	9,09	31,82	59,09	0	0	0	0	0
	/o:/	0	15,00	30	55,00	0	0	0	0	0
	/u:/	0	42,86	14,29	42,86	0	0	0	0	0
	/ø:/	0	0	0	0	0	4,55	9,09	18,18	68,18
	/y:/	0	0	0	0	0	65,22	0	21,74	13,04
	/ɛ/	0	0	0	0	0	0	66,67	14,29	19,05
	/e:/	0	0	0	0	0	0	4,17	37,50	58,33
	/i:/	0	0	0	0	0	0	0	33,33	66,67

Fel kell hívnunk a figyelmet egy mind a 3. táblázatban, mind pedig a 11. ábrán feltűnő, és a korábbi szakirodalmi tételekben nem említett tendenciára. Eszerint azon hangzók esetében, melyek második vagy harmadik nyíltsági fokúak (azaz középső vagy alsó nyelvvállásúak, t.i. az /o: ø: e: ɛ/), tehát egy fokkal nyíltabb és egy fokkal zártabb szomszédjuk is van a magyar magánhangzó-rendszerben, gyakori, hogy az adott ejteni szándékozott magánhangzót nem az ejtési szándék szerinti nyíltsági fokú magánhangzóként azonosították a tesztelők, hanem annál zártabbként. Annak szemléltetésére, hogy ez pontosan milyen mértékben jellemző tendencia, egy összesítő ábrát készítettünk az egyes alaphangfrekvenciákra a válaszok megoszlásáról a magánhangzók nyíltsági fokának tekintetében (12. ábra).



11. ábra: Az ejteni szándékozott magánhangzókra adott válaszok az alapfrekvencia függvényében



12. ábra: A második és harmadik nyíltsági fokú /ε e: ø: o:/ hangzókra az egyes alapfrekvenciákon kapott válaszok megoszlása a magánhangzók nyíltsági fokának függvényében

Az énekelt magánhangzók észlelésének korábbi, az artikuláció és akusztikum, valamint a percepció lineáris összefüggéseit feltételező szakirodalma szerint az alapfrekvencia emelésével megjelenő  $F1 : f_0$  hangolás miatt a magánhangzók az ejteni szándékozott minőségükhöz képest nyíltabbnak valósulnak meg, aminek következtében a percepció rendszer is nyíltabb magánhangzóminőségekként ismeri fel ezeket a hangzókat. A 12. ábrán közölt adatok e felvetés a lehetőségek szerint legjobban kontrollált ellenőrzését teszik lehetővé, ugyanis itt csak azon hangzókat vizsgáljuk, melyek (náluk nyíltabb és zártabb szomszédokkal is rendelkezve) mind nyíltabb, mind pedig zártabb magánhangzókként is azonosíthatóak lehetnek, annak megfelelően, hogy az alapfrekvencia emelése következtében megváltozott minőségük inkább a nyíltabb vagy inkább a zártabb szomszédjukhoz teszi őket hasonlatossá (megjegyzendő, hogy korábban ilyen összevetést egyetlen vizsgálatban sem végeztek).

A 12. ábrán az látható, hogy az egyes alapfrekvenciákon a kiemelt /ε e: ø: o:/ hangzókra adott válaszok miként oszlanak meg – az összevetés szempontja az ejteni szándékozott magánhangzó és a válaszként megjelölt magánhangzó nyíltsága (így az ejtési szándéktól eltérő azonosítás esetén a válasz az ejteni szándékozott hangzónál nyíltabb vagy zártabb, esetleg ugyanolyan nyelvéllásfokú lehet). Az adatok szerint egyértelműen megállapítható, hogy az ejtési szándék szerint második és harmadik nyíltsági fokú (középső és alsó nyelvéllású) magánhangzók esetében az adat-

közlők legnagyobbbrészt az ejteni szándékozott magánhangzónál zártabb magánhangzókat azonosítottak, amikor nem az ejteni kívánt magánhangzókat azonosították a kísérlet során. Ugyanakkor az is látható, hogy ez a tendencia az  $f_0$  emelésével nem mutat egyértelmű összefüggést, egyaránt domináns mindhárom énekelt alapfrekvencián.

Az ejtési szándéktól eltérő azonosítások összesítésekor az is kiderült, hogy az összes stimulus-válasz eltérés között az /i:/ volt a leggyakoribb, ezt a hangzót az esetek 28%-ában jelölték a hallgatók az ejteni szándékozott magánhangzók helyett. A hierarchiában az /i:/-t az /ɒ/ (16%), az /u:/ (14%), az /e:/ (14%), az /o:/ (13%) és az /y:/ (11%) követte.

#### 2.1.2.2. Akusztikai elemzés

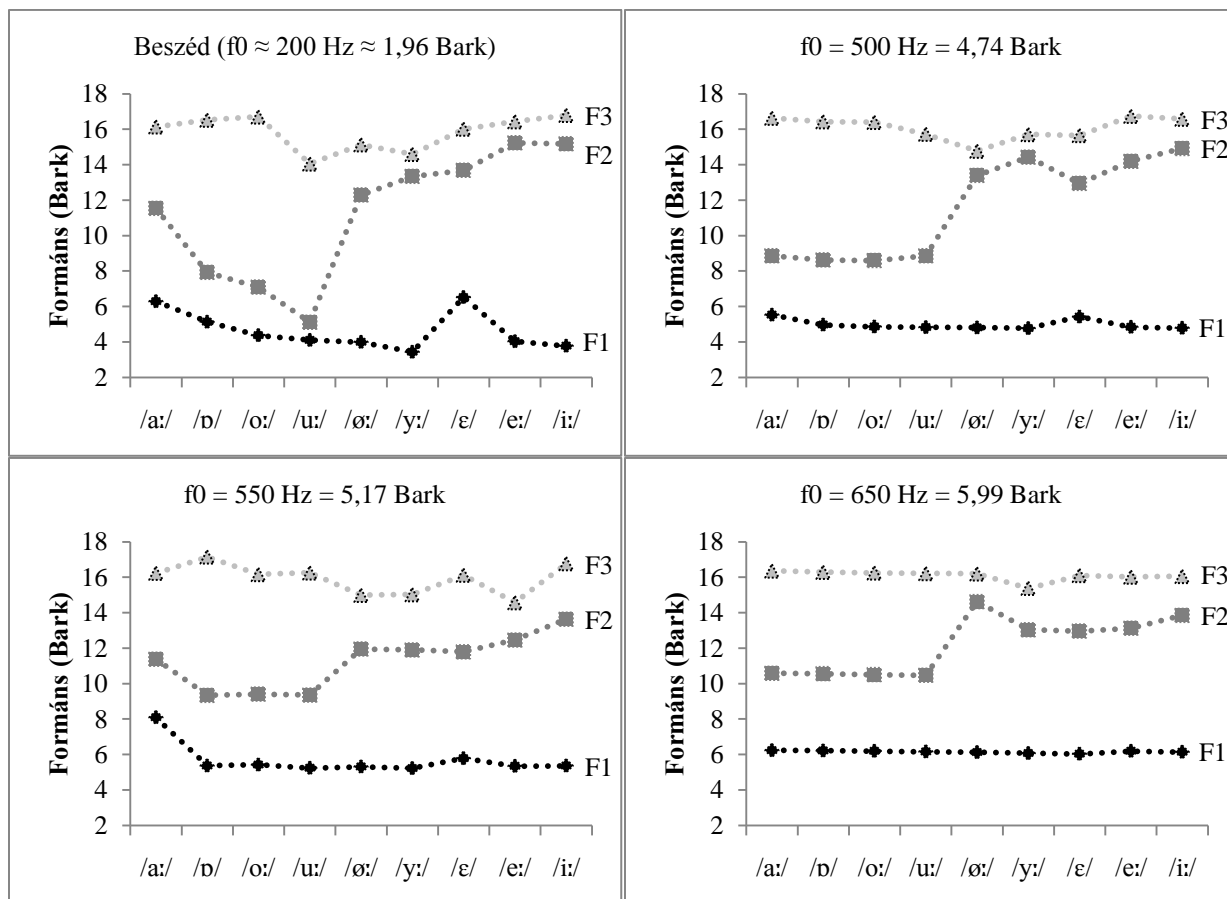
A hangnyomás-idő függvény Fourier-elemzésén alapuló akusztikai méréseink során megállapítottuk a szoprán énekesnek a beszéd módban és ének módban ejtett magánhangzóinak  $F_1$ -,  $F_2$ - és  $F_3$ -értékét. Ezeket érzetileg transzformált, Bark-értékekben<sup>18</sup> kifejezve a 13. ábrán közöljük.

Emlékeztetőül, amint azt már a bevezetésben tárgyaltuk, a formáns fogalom meghatározása alapvetően kettős: míg a kutatók egy része az akusztikai kimenetben mérhető spektrális maximumokat érti rajta, a kutatók egy másik csoportja a toldalékcső mint rezonátorüreg sajátfrekvenciáit, rezonanciáit definiálja (vö. 1.1.1.2. *alfejezet*).

A jelen vizsgálatban kizárólag a mikrofon által rögzített hangnyomás-idő függvényt elemeztük, annak érdekében, hogy az akusztikai jelben mérhető, a percepció rendszer számára elérhető spektrális energiamaximumokat megállapíthassuk. Ebből következően a 13. ábrán látható formánsadatok alacsonyabb alapfrekvencián (de elsősorban a beszédben) nagy valószínűséggel egybeesnek ugyan az artikulációs konfigurációra utaló első három rezonancia frekvenciájával, magasabb alapfrekvenciákon azonban az ábrán közölt adatok nem feltétlenül az egyes üregi rezonanciák középértékét mutatják, mint inkább az adott rezonancia sávszélességébe tartozó legintenzívebb felharmonikus értékét. Éppen ezért ezen adatok alapján az artikulációra vonatkozó következtetések csak megszorításokkal tehetők, az adatok sokkal inkább „csak” a percepció számára elérhető akusztikus prominenciákat mutatják megbízhatóan.

---

<sup>18</sup> A lineáris Hertz értékek ( $f$ ) érzeti transzformációját (*Bark*-ba) Zwicker–Terhard (1980) alapján a következő képlet szerint hajtottuk végre:  $Bark = 13 \times \arctan(0,00076 \times f) + 3,5 \times \arctan((f/7500)^2)$ .



13. ábra: Az énekelve és beszélve ejtett magánhangzók spektrumában mérhető első három formáns mint spektrális energiamaximum értéke (F1, F2, F3)

Akusztikai adataink szerint azon esetekben, amikor az alaphfrekvencia emelése során az énekes hangjának alaphfrekvenciája megközelítette a magánhangzó a beszédben mérhető F1 formánsának értékét, az F1 a megemelt (magas)  $f_0$  értékére vagy afölé hangolódott. Erre abból következtethetünk, hogy az  $f_0$  és F1 az adott (500 Hz) alaphfrekvenciától kezdve nem volt elkülöníthető a magánhangzó spektrumában, miközben a spektrum legintenzívebb akusztikai összetevője az első harmonikus, azaz az  $f_0$  volt (az F1 feltehetőleg tehát ezt a komponenst erősítette). Ennek következményeként az alaphfrekvencia emelésének hatására az F1 értéke uniformizálódott az egyes magánhangzók között, a magánhangzók hangszíne tehát homogenizálódni kezdett a veláris és palatális csoportokon belül. Ez az eredmény egybeesik a korábban bemutatott percepciósi adatokkal, melyek szerint a magánhangzók nyíltsági fokának az ejtési szándék szerinti azonosítása az

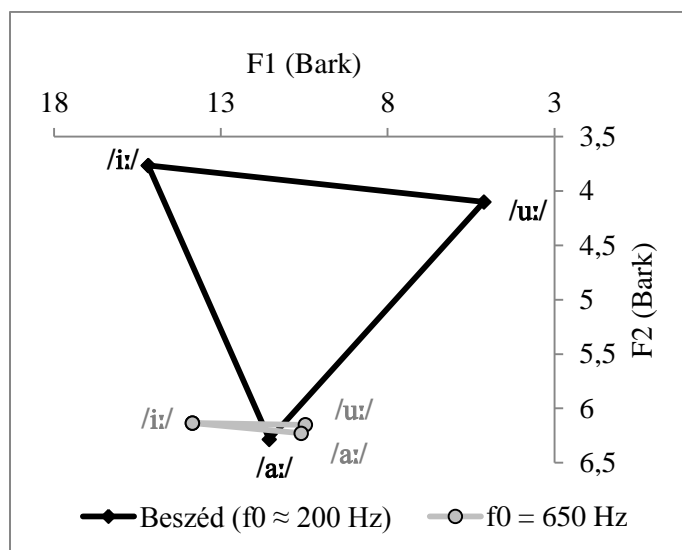
alapfrekvencia emelésével csökken – hiszen az akusztikai eredmények is a nyíltságészlelettel szoros összefüggésben lévő F1 formáns (vagy első spektrális maximum) eltolódását mutatják az alapfrekvencia emelésével. Az F1 emelkedése azonban percepciós adataink szerint nem a magánhangzók észlelt nyíltsági fokának növekedését idézte elő (amint azt korábbi vizsgálatok leírásában feltételezik), hanem sokkal inkább csak a válaszok nagyobb szóródását, tehát a magánhangzó-minőség elbizonytalanodását, amivel együtt az egyes magánhangzók az ejtési szándék szerinti nyíltsági fokánál zártabb magánhangzók azonosítása volt jellemző gyakorlatilag az összes énekelt alapfrekvencián.

Az F2 értékek (azaz a második spektrális maximumhely) alakulásából úgy látszik, hogy az elől és hátul képzett hangzók közötti spektrális különbségek még a kísérletben vizsgált legmagasabb (650 Hz) alapfrekvencián is megmaradtak – mindannak ellenére, hogy a fokozódó spektrális alul-mintavételezettség (vö. 15. ábra) miatt az F2 mint spektrális maximum értékét a legmagasabb alapfrekvencián már gyakorlatilag csak egyetlen felharmonikus határozta meg. A percepciós adatok ezzel az eredménnyel is egybecsengnek: a 11. ábrán bemutatott adatok szerint ugyanis gyakorlatilag egyáltalán nem fordult elő az észlelési tesztben az elől- vagy hátulképzettségnek az ejtési szándéktól eltérő felismerése.

Végezetül összevetettük az artikulációs és akusztikai tekintetben egymástól legmesszebb elhelyezkedő /a: i: u:/ hangok által közrezárt (azok spektrális maximumai által meghatározható) akusztikai magánhangzóteret a beszédben és a kísérletben legmagasabb énekelt alapfrekvencián (14. ábra).

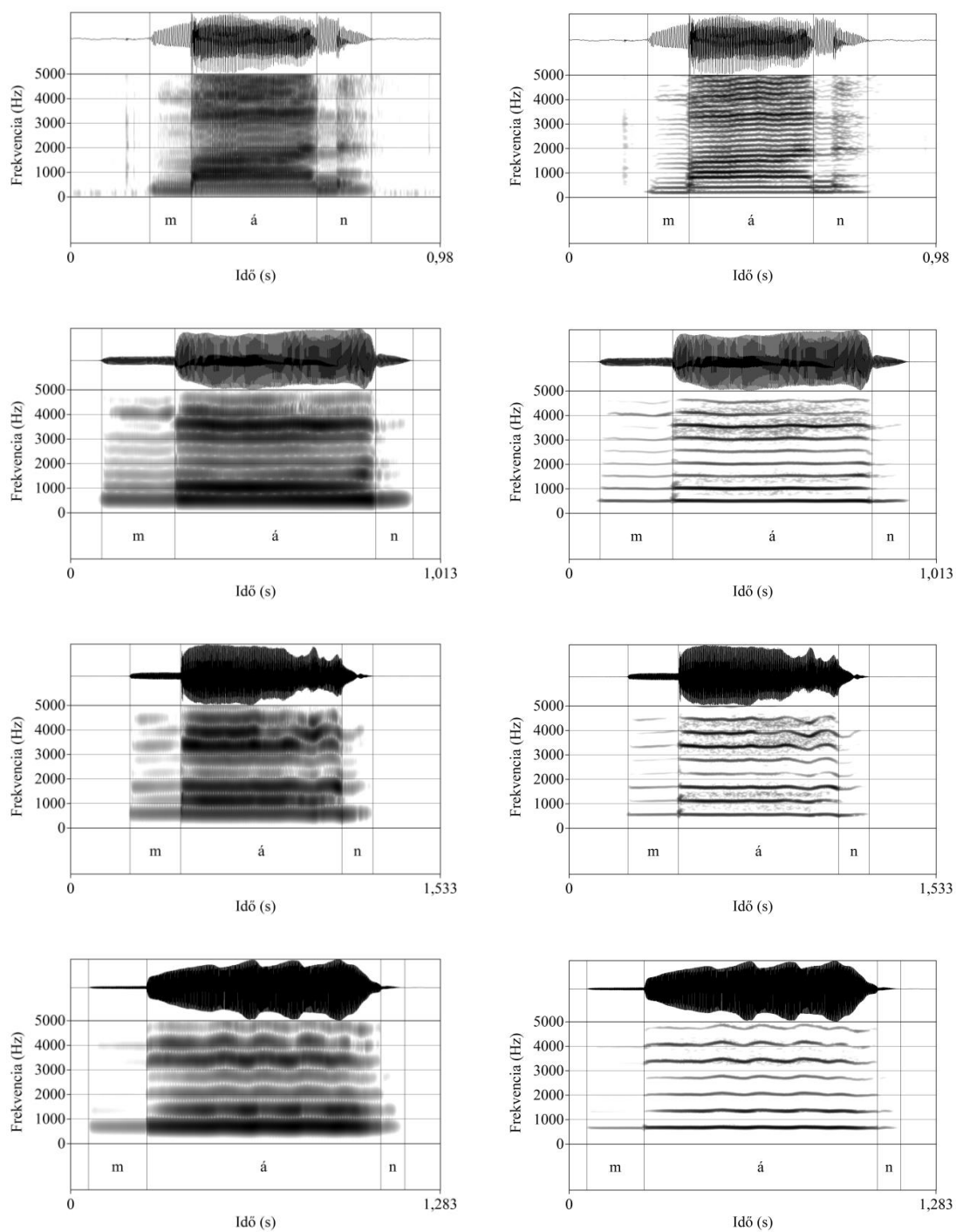
A várakozásoknak megfelelően az akusztikai magánhangzóter az alapfrekvencia emelésével jelentősen redukálódott, és eltolódott az /a:/ pozíciója felé, ugyanakkor jól láthatóan 650 Hz alapfrekvencián az elől- és hátulképzettség dimenziója nem csökken le olyan mértékben, mint a nyíltságnak szokásosan megfeleltethető F1 dimenzió.





14. ábra: Az akusztikai tekintetben egymástól legmesszebb elhelyezkedő /a: i: u:/ magánhangzók által meghatározott akusztikai magánhangzótér a beszédben és a kísérlletben legmagasabb énekelt alapfrekvencián

A 15. ábrán a *mán* hangsor hullámformája, valamint szélessávú (bal oldal) és keskenysávú (jobb oldal) spektrogramja látható. Az első sorban a beszédbeli megvalósítást találjuk, a másodiktól negyedik sorban pedig (rendre) az 500, 550 és 650 Hz alapfrekvenciákon énekelt hangsorokat. A spektrogramokon (de különösen a keskenysávú spektrogramon) jól kivehetők a harmonikusok (az alaphang és felharmonikusai, melyek vízszintes csíkokként jelennek meg), így az is megfigyelhető, hogy az alapfrekvencia emelésével mint kerülnek egymástól egyre nagyobb és nagyobb távolságra. A felhangok intenzitására ebben a megjelenítésben a sötétedés mértéke utal – így tehát a toldalékcso rezonanciái (legalábbis a beszédjelben) sötétebb sávokként jelennek meg a spektrogramon. Jól látható, hogy míg a beszédjelben a felerősített felharmonikus-nyalábok egyértelműen kirajzolják az egyes (sötétebb sávokként megjelenő) rezonanciahelyeket, azaz a formánsokat, addig a magasabb alapfrekvenciákon a formánsok frekvenciahelyének detekciója már sokkal bizonytalanabb. Ezeken az alapfrekvenciákon az intenzitásérősödések már sokkal inkább csak az egyes felharmonikusok erősítettségére utalnak megbízhatóan, a formánsok középértékére azonban nem – „formánsértékeként” az akusztikai kimenetből tehát csak az erősített felharmonikusok frekvenciaértékei, nem pedig az üregi rezonanciák középértéke azonosítható.



15. ábra: A beszédben ejtett (első sor), valamint az 500, 600 és 650 Hz alapfrekvencián énekelt (2–4. sor) *mán* hangsor hullámformája (a panelek felső sorában), valamint széles-sávú (bal oldal) és keskenysávú (jobb oldal) hangszíneképe

### 2.1.3. Következtetések

Az itt ismertetett vizsgálat az első olyan kísérletsorozat, mely mind produkciós, mind pedig percepciós szempontból igyekezett feltérképezni az éneklés mint produkciós mód hatását egy adott nyelv összes magánhangzójára. Eredményeinket, és főként az észlelésben tapasztalt tendenciákat a magyar magánhangzók rendszerének figyelembevételével értékeltük, mely szempont az eddigi vizsgálatokhoz képest szintén újításnak számít.

Percepciós vizsgálatunk eredményeit összegezve megállapítható, hogy kísérletünkben nem igazolódott az a szakirodalomban olvasható (de alapvetően csak spekulatív) feltételezés, mely szerint az alapfrekvencia emelésével (az  $F1 : f_0$  hangolásának következtében) egyre magasabb  $F1$  a megvalósult magánhangzók percepciós nyitódását okozná. Ezt az állítást arra a jól kontrollált összevetésre alapozhatjuk, amelyben a nyíltabb és zártabb szomszédokkal rendelkező második és harmadik nyíltsági fokú magánhangzókra kapott válaszokat elemeztük. Eszerint ugyanis magas alapfrekvenciákon (már a kísérletben legalacsonyabb énekelt alapfrekvenciaként vizsgált 500 Hz fölött) a legjellemzőbb tendencia az egyes magánhangzók az ejtési szándéktól eltérő azonosításában a zártabb hangzóként való felismerés. A korábbi vizsgálatokban azonban vagy nem derül ki pontosan, mire alapozták a szerzők megállapításukat vagy feltételezéseiket a hangzók nyíltabbaként történt felismeréséről (vö. Scotto di Carlo–Germain 1985), vagy csak olyan zárt (első nyíltsági fokú) magánhangzók vizsgáltak a szerzők, melyek – nem lévén zártabb szomszédjuk – „tévesztés” esetén eleve csak nyíltabb hangzókként voltak azonosíthatók (vö. Hollien et al. 2000). Így azokat a jelen vizsgálat eredményeivel nem volna érdemes ütköztetnünk. Az ejtési szándéktól eltérő azonosításokról, valamint az alapfrekvencia emelésével megjelenő észlelési tendenciákról (általánosan) annyit mondhatunk, hogy az  $f_0$  emelésével egyre csökken az egyes magánhangzók az ejtési szándék szerinti azonosításának aránya, miközben a válaszok szóródása egyre növekszik – ugyanis a válaszként adott magánhangzók egyre nagyobb szóródást mutatnak a különböző magánhangzó-minőségek között.

A kísérletben nem igazolódott az a korábban megfogalmazott feltételezés sem, hogy a magasabb alapfrekvenciákon a nyíltabb magánhangzók felismerése volna a legjobb. Bár a kísérletben elért 650 Hz alapfrekvenciát a szoprán hangterjedelem szempontjából nem tekinthetjük kiemelkedően magasnak (a szoprán hangterjedelmet felülről átlagosan a „magas C” vagy a  $c'''$ , azaz

1047 Hz alaphfrekvencia határolja, l. pl. McKinney 1994), mégis ezen a – mondjuk így – közepesen magas alaphfrekvencián a legmagasabb az ejtési szándék szerinti azonosítást a zárt, első nyíltsági fokú /i:/ és /y:/ mutatta, valamint a harmadik nyíltsági fokú /e/. Ezzel párhuzamosan az /i:/ más szempontok szerint is kiemelt szerepét láttuk: az /i:/ volt ugyanis a leggyakoribb válasz az ejtési szándéktól eltérő azonosítások során.

Az /i:/ különleges percepció tulajdonságait már több vizsgálatban is kimutatták. Az éneklésre Scotto di Carlo és Germain (1985) jegyezte le először az /i:/ mint gyakori válasz megjelenését a magasabb alaphfrekvenciákon, a beszédre Mády és Reichel (2007) is kimutatta azt az összefüggést, mely szerint a hosszan ejtett /e:/ magánhangzóra gyakran a nála zártabb minőségű /i:/ válasz érkezett (míg más magánhangzók hosszú ejtésénél nem volt jellemző a hangzók zártabbként való azonosítása). A Scotto di Carlo és Germain (1985) által csak a korábbi artikulációs vizsgálatok alapján feltételezett, Mády és Reichel (2007) által pedig akusztikai mérések alapján megállapított formánsadatok egyik esetben sem magyarázták ezeket a „tévesztéseket”. Szintén említhető itt Gósynak (1989) a magánhangzók azonosítását vizsgáló kísérlete, melyben a szerző különböző szűrők alkalmazásával nehezítette a magánhangzók a percepció számára fontosnak vélt akusztikai kulcsainak percepció elérését. Ebben a vizsgálatban ugyanis azt találták, hogy az /i:/ és /y:/ magánhangzók azonosítása a csak igen magas frekvenciájú összetevőket tartalmazó frekvenciasáv (2700–3300 Hz) alapján is viszonylag jó arányú maradt (vö. 1.1.2.3. *alfejezet*). Az /i/ hangzót Ryalls és Liberman – korábbi percepció vizsgálatokra is hivatkozva – „szupermagánhangzónak” nevezi (1982: 1633), mégpedig azért, mert az összes általuk vizsgált magánhangzó közül az /i/-t találták a legkevésbé érzékenynek a különböző „nehezített” percepció feladatokban. Magyarázatauk szerint az /i/ kitüntetett szerepét különös szalienciájának köszönheti, distinktív akusztikai jegei ugyanis minden más magánhangzónál jobban elkülöníthetővé teszik a magánhangzórendszer (vagy talán inkább a magánhangzótér) többi tagjától. Akárhogy is (talán ezekkel a korábbi eredményekkel összefüggésben, talán nem), a jelen percepció vizsgálatban igen erős /i:/ preferenciát találtunk a válaszadásban, melynek magyarázata pusztán az egyes énekelt magánhangzók megvalósulásának akusztikai vetülete (és az F1 és F2 abszolút értéke) alapján nem tűnik lehetségesnek.

A magas alaphfrekvencián énekelt magánhangzók akusztikai vetületében a magas alaphfrekvencia (és a magas alaphfrekvencián ejtett magánhangzók spektrális alul-mintavételezettsége) miatt

elsősorban nem a toldalékcsőre mint üregrendszerre az artikuláció során jellemző rezonanciák, hanem a megvalósított hangzókban tapasztalható energiamaximumok mérhetők, melyek megfelelően magas  $f_0$  esetén nem egyeznek a toldalékcső rezonanciahelyeivel. A jelen alfejezetben ismertetett mérések során ennek megfelelően nem az artikulációt többé-kevésbé pontosan leíró rezonanciafrekvenciákat, hanem a kisugárzott hang spektrumának energiamaximumait detektáltuk. Ez az információ nagyon fontos, hiszen a beszédészlelés számára rendelkezésre álló prominenciákra utal, ugyanakkor ezekből az adatokból a magánhangzók létrehozásakor jellemző artikulációs működésekre csak nagyon áttételesen következtethetünk.

Ezek alapján az adatok alapján arra következtethetünk, hogy valószínűleg fellép az  $F1 : f_0$  hangolás azokban az esetekben, amikor a megemelt  $f_0$  magasabb lenne, mint az adott hangzónak a beszédben jellemző első formánsa. Ezt főként arra a megfigyelésre alapozzuk, hogy magasabb alapfrekvenciákon minden esetben az  $f_0$  volt a legintenzívebb frekvenciakomponens, melytől az  $F1$  elkülöníthetetlennek bizonyult. Az  $F2$  ezzel szemben a 650 Hz alapfrekvencia eléréséig nem uniformizálódott teljesen a magánhangzók között (még annak ellenére sem, hogy a spektrális alulmintavételezés miatt csökkent az  $F2$ -höz mint rezonanciához eső spektrális maximumot kialakító felharmonikusok száma). Így a palatális és veláris magánhangzócsoportok az  $F2$  mentén mind akusztikailag, mind (az észlelési teszt alapján) perceptuálisan elkülönültek egymástól még ezen az alapfrekvencián is.

Az akusztikai magánhangzótér az  $F1$  és  $F2$  formáns változásának következtében a legmagasabb alapfrekvenciához (a beszédhez képest) jelentősen redukálódott, és a nyílt, centrális /a/ pozíciója felé tolódott. Érdekes kérdés, hogy ennek ellenére miért nem növekedett meg drasztikusan az /a:/ jelölések száma a válaszok között. A kérdés megválaszolásához azonban feltehetőleg több kell, mint az egyes megvalósított hangzók első két spektrális prominenciájának itt mért akusztikai adatai, valamint annak a feltételezése, hogy az artikuláció és az akusztikum a percepcióval egyszerű, lineáris összefüggésben áll. Mindezek helyett valószínűleg gyümölcsözőbb volna azokat a percepciók eredményeket és modelleket segítségül hívni az itt tapasztalt észlelési folyamatok megértéséhez, melyek a zártságot inkább az  $f_0$  és  $F1$  tonotópiás viszonyának függvényében értelmezik. Traunmüller (1981) eredményei szerint a nyíltság érzeti dimenziójában az  $F1$  abszolút frekvenciaértéke helyett sokkal inkább az  $f_0$  és  $F1$  távolsága a döntő (vö. a jelen értekezés

*1.1.2.2. alfejezete*), mely az éneklésben az alapfrekvencia emelésével minden bizonnyal változó paraméter. Az  $F1 : f_0$  hangolás megléte vagy hiánya azonban az elmélet alkalmazása szempontjából kritikus kérdés, ahogyan az  $F2$  értéke is – az elmélet alkalmazásához ugyanis szükséges lenne (a jelen adatoknál megbízhatóbban) tudnunk, hogy pontosan mely formánsok, és mely felharmonikusokat erősítik. Az alfejezetben bemutatott adatok alapján ezért (egyelőre) kérdéses marad, miként volna adaptálható a modell az éneklésben tapasztalt észlelési tendenciákra; és csak sejthető, hogy az  $F1-f_0$  csökkenő tonotópiás távolsága lehet a felelős a magánhangzóknak az ejtési szándéknál zártabb észleletére. A későbbiekben azonban, a *2.5. alfejezetben* bemutatott kísérlet során – melyben rezonanciaadatokat és észlelési adatokat egyaránt vizsgálni fogunk –, még ennél részletetbe menőbb módon is megkíséreljük az elmélet adaptációját. Addig is (a jelen vizsgálat fontos konklúziójaként) kiemelendő az a jelen vizsgálatból nyert egyértelmű empirikus bizonyíték, mely szerint a pontosan kontrollált és releváns összehasonlításokat tartalmazó percepciós vizsgálat nem utal a produkció és az észlelés közötti egyszerű, lineáris összefüggésekre, melyeket korábbi vizsgálatok során egyértelműen elfogadhatónak tartottak.

Abból, hogy az alapfrekvencia emelésével egyre csökkent a magánhangzóknak az ejtési szándék szerinti azonosítása, valamint abból, hogy a válaszok szóródása az egyre magasabb  $f_0$  elérésével egyre nagyobb lett az következik, hogy az énekelt magánhangzók akusztikai és percepciós minősége az alapfrekvencia emelésével együtt változott. E változás természete az eredmények szerint nem kategoriális, azaz az adatközlők nem egy észleleti kategóriáról hirtelen a másikra váltva azonosították az egyes hangzókat, hanem fokozatos, aminek következtében a magánhangzók percepciós minősége (legalábbis 650 Hz alapfrekvenciáig) az alapfrekvencia emelésével egyre bizonytalanabbnak tűnt fel. Mindezek fényében (az *1. fejezet* végén megfogalmazott hipotéziseinkre is utalva) úgy véljük, kérdéses, hogy várható-e bármilyen pozitív hatás a mássalhangzó-környezettől mint a koartikulációs formánsátmenetek révén a közbeeső magánhangzó minőségét kulcsoló percepciós „mankóktól”. Ha ugyanis a magánhangzók minősége valóban megváltozik, a formánsátmenetek is legfeljebb csak erre a megváltozott magánhangzó-minőségre (nem pedig egy „elméleti”, de meg nem valósuló magánhangzó-minőségre) utaló kulcsokat tartalmazhatnak. A kérdésre a *2.4. alfejezetben* bemutatott kísérletekkel keressük a választ.

További (részben az előzőhöz kapcsolódó) kérdésként merül fel, hogy vajon hatással volt-e a jelen kísérletben használt mássalhangzó-környezet minősége, azaz a két nazális mássalhangzó az észlelési eredményekre. Az internazális magánhangzó nazalizációja ugyanis megváltoztatja a magánhangzók akusztikai szerkezetét, az akusztikai jellemzők között is leginkább az F1 értékét (Krakow et al. 1987 és hivatkozásai, Rosner–Pickering 1994, Horváth 2005, Horváth 2008). Ez alapján tehát logikusan feltételezhető lenne, hogy a nazalizáció a magánhangzók észlelt nyíltságára is hatást gyakorolhat a beszédben. Ugyanakkor egyes vizsgálatok szerint e hatást, ha természetes, tehát bejósolható helyen fordul elő a beszédben (azaz egy magánhangzó nazális, nem pedig orális mássalhangzók mellett/között nazalizálódik), a percepció rendszer képes kompenzálni, így nem változik a magánhangzó zártságának észlelete (Beddor 1983, Krakow et al. 1987). Vitatott kérdés, hogy az énekelt nazális mássalhangzók a beszédben ejtettekkel egyező mértékben nazalizáltak-e, így nincs egyértelmű állásfoglalás abban a kérdésben sem, hogy az éneklésben megvalósuló nazális mássalhangzók hatása a beszédben tapasztalhatóhoz képest milyen mértékű nazalizációt fejt ki a szomszédos magánhangzókra (vö. Vennard 1964, Scotto di Carlo–Germain 1985, Yangisawa et al. 1990, Yangisawa et al. 1991, Gramming et al. 1993). Éppen ezért fontosnak tartjuk megvizsgálni ezt az aspektust is, és kizárni annak a lehetőségét, hogy a jelen vizsgálat eredményeit a magánhangzók nazalizációja eredményezte volna.

Az itt felvetett problémák, azaz a nazális mássalhangzó-környezet hatásának ellenőrzésére, valamint a mássalhangzó-környezetnek az izolált ejtéssel összehasonlításban tapasztalható hatásának vizsgálatára terveztük a 2.2. és 2.3. *alfejezetekben* bemutatásra kerülő kísérleteket. A 2.5. *alfejezetben* pedig, amint azt már említettük, a formánsok mint üregi rezonanciák vizsgálatát kíséreljük meg, és ezeket az adatokat észlelési vizsgálatok eredményeivel is összevetjük.

## **2.2. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: a mássalhangzó képzésmódjának hatása**

Előző kísérletünkben a magyar magánhangzók akusztikai és percepció tendenciáit vizsgáltuk az éneklésben, mint különleges produkciós sajátosságokkal rendelkező produkciós módban. Az

észlelési eredmények szerint az alapfrekvencia emelésével a magánhangzók az ejtési szándék szerinti azonosításának aránya csökkent, miközben olyan az ejtési szándéktól eltérő azonosítási tendenciákat találtunk, melyek nem támasztják alá azokat a korábbi szakirodalomban sugallt feltevézéseket, amelyek a produkció és percepció közti egyszerű, lineáris összefüggéseken alapulnak.

Annak ellenőrzésére, hogy eredményeinket valóban elsősorban az éneklés mint sajátos artikulációs és akusztikai jellemzőkkel bíró produkciós mód hatására kaptuk, és nem más körülmények, például a választott nazális mássalhangzó-környezet miatt, a jelen vizsgálatban kibővítjük a korábbi észlelési tesztünket. A nazális hangkörnyezet mellett ez alkalommal zöngés és zöngétlen frikatívákat is tesztelünk.

Az orális magánhangzók nazális mássalhangzók környezetében a koartikuláció következtében nazalizálódhatnak, azaz a mássalhangzó képzésére jellemző nazális jelleg áterjedhet a szomszédos magánhangzóra. A nazalizáció eredményeként az eredetileg orális minőségű magánhangzók minősége megváltozik, a szakirodalom szerint a legjelentősebb változás az F1 formánst érinti, mégpedig a következőképpen. Korábbi akusztikai mérések szerint a nazalizáció hatására a felső nyelvállású vagy első nyíltsági fokú magánhangzók, valamint egyes második vagy harmadik nyíltsági fokú magánhangzók F1-értéke nő, azaz a magánhangzók akusztikailag nyílnak, míg a többi második és harmadik nyíltsági fokú magánhangzó, valamint a negyedik nyíltsági fokú magánhangzók F1-értéke csökken, azaz ezek a magánhangzók akusztikailag záródnak. Mindeközben a nazalizált magánhangzó első spektrális prominenciája minden esetben nagyobb sávszélességű és kisebb intenzitású, mint az orális realizációé (vö. Krakow et al. 1987).

Az akusztikai változások következményeként sok kutató azt feltételezte, hogy a nazalizált magánhangzók az orális magánhangzókként megvalósított megfelelőjükhöz képest kevésbé felismerhetők a beszédben. Erre például Wright (1975) szolgáltatott az elsők között kísérletes bizonyítékokat izolált ejtésű nazalizált magánhangzók vizsgálatával. Később azonban Krakow és munkatársai (1987) arra is fényt derítettek, hogy a nazalizáció csak a hangkörnyezetre vonatkozó információk hiányában hat negatívan az ejtés szerinti azonosításra. A szerzők ugyanis azt találták, hogy a nazális mássalhangzók között ejtett magánhangzók azonosítása az orális mássalhangzók között ejtett magánhangzókéhoz nagyon hasonlóan alakult. Ebből Krakow és munkatársai (1987)



arra következtettek, hogy bár a nazalizáció megváltoztatja a magánhangzó akusztikai szerkezetét, hatása bejósolható helyen, azaz nazális mássalhangzók környezetében jelentkezve perceptiósan „korrigálható”.

Mivel a megfelelő technikával képzett énekhang sok énekes tapasztalata szerint orrüregi rezonanciaélménnyel jár (erre utal az „előre helyezett”, „elől képzett” hang gyakran használt metaforája is), az éneklés szakirodalmában sokat kutattak, de máig vitatott kérdés a nazalizáció megjelenése és a nazalizáció mértéke az éneklésben. Egyes kísérletek tanúsága szerint az éneklésben a nazális minőségűnek szánt mássalhangzók nazalizációjának mértéke és a szomszédos magánhangzóra kifejtett nazalizációs hatás – főként magasabb alapfrekvenciákon – jóval kisebb, mint a beszédben, sőt akár teljesen el is tűnik (Vennard 1964, Yangisawa et al. 1990, Yangisawa et al. 1991), mások szerint azonban a lágy szájpad az éneklésben valamelyest folyamatosan nyitva van, ráadásul úgy, hogy ebben az artikulációs stratégiában nagy egyéni variabilitás tapasztalható, miközben az orrüregi csatolás mértéke is igen változatos lehet, mind az adott magánhangzó, mind a beszélő személy függvényében (Gramming et al. 1993). Mindemellett fontos azt is megjegyeznünk, hogy az énekelt magánhangzók nazalizációjának perceptiós vetületéről a jelen értekezés szerzőjének tudomása szerint még nem született vizsgálat, nem tudjuk tehát, hogy ha van is orrüregi csatolás az éneklés során, az egyáltalán érzékelhető-e bármilyen módon a hallgatók számára.

A fentiekből következően nem zárható ki tehát teljesen, hogy az előző kísérletünkben vizsgált magánhangzók megvalósulását befolyásolta a nazális hatás, így a perceptiójuk – valamilyen formában – szintén a nazalizáció hatása alatt állt. Éppen ezért a jelen alfejezetben bemutatott vizsgálat célja e kérdés ellenőrzése volt. Ehhez nazális mássalhangzó-környezetben ejtett magánhangzók azonosítását vetettük össze zöngés és zöngétlen posztalveoláris frikatívák között ejtett magánhangzók azonosításával. Mivel a szakirodalom alapján tudjuk, hogy a nazalizált magánhangzók nazális hangkörnyezetben gyakorlatilag azonos százalékbán azonosíthatók, mint az obstruens környezetben ejtett megfelelőjük (vö. Krakow et al. 1987), a vizsgálat elején azt feltételeztük, hogy a nazális és frikatíva hangkörnyezetek hatása között nem találunk jelentős különbségeket az éneklésben sem, mert ha igaz is, hogy az éneklésben a nazalizációs hatás megjelenik, az a ma-

gánhangzók hangkörnyezetben történő bemutatásakor nem változtat jelentősen a magánhangzó észlelt hangszínén.

### 2.2.1. Anyag, módszer és kísérleti személyek

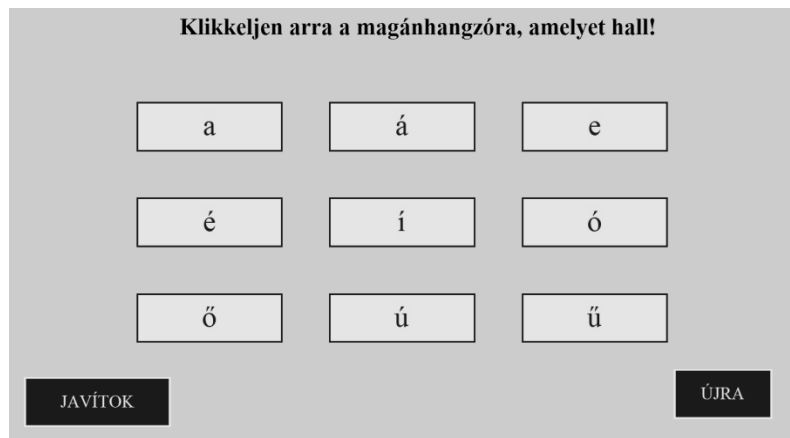
Az észlelési kísérlethez az előző vizsgálatban részt vevő szoprán énekesnővel (életkora 50 év) készítettünk felvételeket (csendesített szobában, Zoom H4 típusú irányított mikrofonnal, 44,1 kHz-en, 16 biten). Az énekes ismét az /*ɒ a: ɛ e: i: o: ø: u: y:/* magánhangzókat ejtette ki beszéd módban ( $\approx 200$  Hz alapfrekvencián), valamint ének módban, utóbbi esetben az 500, 550, 600, 650 Hz alapfrekvenciákon, statikus dallamívvel, kitartva. Az énekelt stimulusok felvételekor ismételten a megfelelő *f0* értéken megszólaló referenciahangokkal segítettük a pontos intonációt. Minden hangsort körülbelül 1,5–2 másodpercig hangoztatott az énekes. A magánhangzókat ez alkalommal az *mVn* valamint az *sVs* és a *zsVzs* hangkörnyezetekben vettük fel (pl.: *mán, sás, zsászs* stb.). Így összesen ( $3 \text{ kontextus} \times 9 \text{ magánhangzó} \times 5 \text{ alapfrekvencia} \times 2 \text{ ismétlés} =$ ) 270 stimulust teszteltünk disztraktorokkal kiegészítve (a disztraktorok más hangkörnyezetben ejtett magánhangzók voltak).

Az észlelési tesztben 15 naiv (tehát sem a beszédtechnikában sem az éneklésben nem képzett) adatközlő vett részt (10 nő és 5 férfi, 21-29 évesek, átlagos életkoruk 23 év). A percepciók tesztet egy erre a célra a Praat programban (Boersma–Weenink 2009) írt szkript vezérelte. Az adatközlők feladata az volt, hogy az egyes szekvenciák elhangzása után kattintsanak a képernyőn helyesírásban bemutatott magánhangzók közül arra, amelyiket a hangsorban hallani vélték (16. ábra). A teszt előtt felhívtuk rá a kísérleti személyek figyelmét, hogy értelmes és értelmetlen egy szótagú szavak egyaránt előfordulhatnak a tesztben, ezért minden esetben a magánhangzó minőségére figyeljenek, ne az elhangzó szót vagy álszót próbálják azonosítani. A résztvevőknek lehetősége volt egyszer újrahallgatni az aktuális hangsort (az *Újra* gomb megnyomásával), valamint egyszer javítani az esetleges félrekattintásból származó eredményt (a *Javítok* gomb megnyomásával). Ez után azonban mindenképpen választaniuk kellett.

A tesztet a résztvevők egyesével végezték el, a stimulusokat fülhallgatón keresztül hallották.

Az észlelési eredményeken statisztikai próbákat végeztünk az SPSS 13.0 programmal: előbb Shapiro–Wilk-próbával elemeztük a normális eloszlás kritériumát, majd ennek teljesülése esetén

egyváltozós varianciaanalízist (ANOVA), nem normális eloszlás esetén pedig nem parametrikus tesztek ( $\chi^2$ -próba, Mann-Whitney U-próba) alkalmaztunk.



16. ábra: A percepció teszt felülete a Praat programban

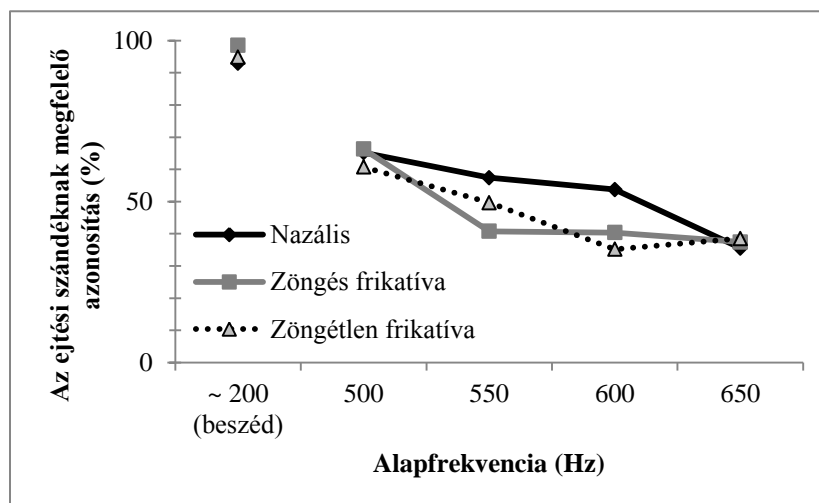
### 2.2.2. Eredmények

A beszédben és az éneklésben tapasztalt, az alapfrekvencia emelésével megjelenő átlagos azonosítási tendenciákat összegzi a 17. ábra. Az eredmények szerint a mássalhangzó-környezet minősége nincs szignifikáns hatással a magánhangzók azonosítására (ANOVA,  $F(2) = 0,065$ ,  $p = 0,937$ ), mégis az látható, hogy a nazális kontextus 550 és 650 Hz alapfrekvencián valamivel magasabb, az ejtési szándéknak megfelelő azonosítási arányokat eredményezett, mint a zöngés és zöngétlen frikatívák, 500 és 600 Hz alapfrekvencián, valamint a beszédben azonban gyakorlatilag nem látunk különbséget.

Az ejtési szándéktól eltérő azonosítási tendenciák megfigyeléséhez a stimulus-válasz párokat a korábbiakhoz hasonlóan ismét konfúziós mátrixokba rendeztük. A legmagasabb alapfrekvencián tapasztalt azonosítási tendenciákat e mátrixokba rendezve a 4. táblázatban közöljük.

A mátrixokban az 50%-os és az afeletti azonosításokat besötétítettük, hogy a kimagaslóan magas arányban az ejtési szándék szerint azonosított magánhangzókat összevethessük az egyes kontextusokban (ehhez az 50%-os határt önkényesen választottuk). Amint az látható, a legmagasabb alapfrekvencián a nazális környezetben az /a:/, /u:/ és /e:/, zöngés frikatívák között az /a:/, /u:/ és /y:/, míg zöngétlen frikatívák között az /y:/ és az /o:/ azonosítása volt a legnagyobb arányban az

ejtési szándéknak megfelelő. Annak a figyelembe vételével, hogy kilenc választási lehetőség esetén annak az esélye, hogy az adatközlők véletlenszerűen választják az ejtési szándék szerinti magánhangzót, 11%, azt is megállapíthatjuk, hogy ezen a kísérletben vizsgált legmagasabb alapfrekvencián a nazális környezetben csak az /ø:/, míg a zöngés frikatívák között csak az /ɒ/ ejtési szándéknak megfelelő azonosítása csökkent a véletlen szintje alá, míg a zöngétlen frikatívák között egyetlen hang azonosítása sem. Bár az ejtési szándéktól eltérő minőségű magánhangzó-azonosításokkor adott válaszok szóródása az alapfrekvencia emelésével egyre nagyobbak bizonyult, a legmagasabb alapfrekvenciára kapott adatokból is látható módon a palatális és veláris magánhangzócsoporthoz még 650 Hz alapfrekvencián is egymástól elkülönülten jelentek meg a percepcióban.



17. ábra: Az énekelt magánhangzók azonosítása nazális, valamint zöngés és zöngétlen frikatíva környezetében az alapfrekvencia emelésével

Ahhoz, hogy feltérképezzük az egyes alapfrekvenciákon jellemző az ejtési szándéktól eltérő minőségű magánhangzók azonosításának tendenciáit, ismét kiemeltük és összesítettük a második és harmadik nyíltsági fokú /o: ø: e: ε/ hangzókra kapott válaszokat az  $f_0$  függvényében (18. ábra).

Bár az egyes hangkörnyezetekre kapott válaszok az magánhangzó észlelt nyíltsága szerint nem teljesen egységesek, jól látható, hogy minden középső vagy alsó nyelvállású, azaz második vagy harmadik nyíltsági fokú magánhangzó (tehát az /ε e: ø: o:/) esetében a magánhangzók az ejtési

szándéknál zártabb azonosítása volt jellemző. Ezen túlmenően pedig azt a korábban (az átlagadatokon) tapasztalt tendenciát is megfigyelhetjük, hogy a nazális hangkörnyezetben a legmagasabb az ejtési szándéknak megfelelő azonosítás gyakorlatilag minden alaphangfajta esetén.

4. táblázat: A kísérletben legmagasabb alaphangfajta (650 Hz) tapasztalt azonosítási tendenciák a három mássalhangzó-környezetben

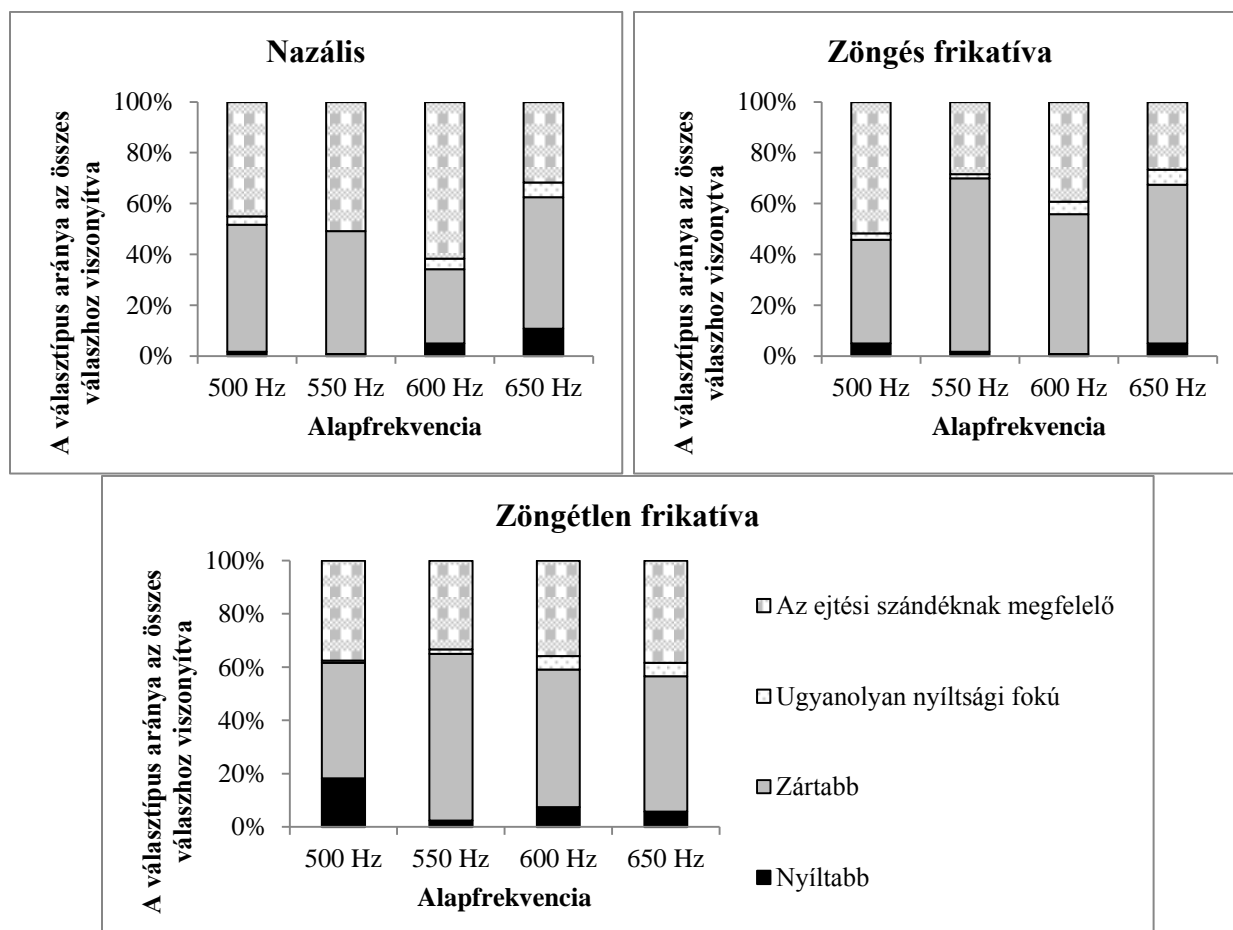
Nazális		Válasz (%)								
		/a:/	/b/	/o:/	/u:/	/ø:/	/y:/	/ɛ:/	/e:/	/i:/
Az ejteni szándékozott magánhangzó	/a:/	63,3	23,3	3,3	10	0	0	0	0	0
	/b/	23,3	30	33,3	13,3	0	0	0	0	0
	/o:/	16,7	13,3	40	30	0	0	0	0	0
	/u:/	3,3	3,3	36,7	56,7	0	0	0	0	0
	/ø:/	3,3	0	0	0	6,7	40	6,7	23,3	20
	/y:/	0	0	0	0	0	26,7	6,7	30	36,7
	/ɛ:/	0	0	0	0	6,7	10	30	43,3	10
	/e:/	0	0	0	0	0	3,3	3,3	50	43,3
	/i:/	0	0	0	0	3,3	3,3	3,3	73,3	16,7

Zöngés frikatíva		Válasz (%)								
		/a:/	/b/	/o:/	/u:/	/ø:/	/y:/	/ɛ:/	/e:/	/i:/
Az ejteni szándékozott magánhangzó	/a:/	76,7	20	0	0	0	0	0	3,3	0
	/b/	6,7	10	63,3	20	0	0	0	0	0
	/o:/	3,3	3,3	33,3	60	0	0	0	0	0
	/u:/	6,7	3,3	26,7	63,3	0	0	0	0	0
	/ø:/	0	0	0	0	20	60	3,3	10	6,7
	/y:/	0	0	0	0	26,7	50	3,3	13,3	6,7
	/ɛ:/	6,7	0	0	0	46,7	10	13,3	20	3,3
	/e:/	0	0	0	0	13,3	20	3,3	40	23,3
	/i:/	0	0	0	0	0	13,3	0	56,7	30

Zöngétlen frikatíva		Válasz (%)								
		/a:/	/ɒ/	/o:/	/u:/	/ø:/	/y:/	/ɛ/	/e:/	/i:/
Az ejtési szándékozott magánhangzó	/a:/	43,3	33,3	20	3,3	0	0	0	0	0
	/ɒ/	6,7	23,3	46,7	23,3	0	0	0	0	0
	/o:/	6,7	13,3	60	20	0	0	0	0	0
	/u:/	3,3	0	56,7	40	0	0	0	0	0
	/ø:/	3,3	0	0	0	20	60	0	10	6,7
	/y:/	0	0	0	0	16,7	73,3	0	6,7	3,3
	/ɛ/	0	0	0	3,3	26,7	16,7	46,7	6,7	0
	/e:/	0	0	0	0	10	33,3	0	26,7	30
	/i:/	0	0	0	0	13,3	63,3	3,3	6,7	13,3

A mássalhangzó-környezeteket összehasonlítottuk a nem az ejtési szándék szerinti azonosítások átlagosan jellemző aránya szerint is, tehát aszerint, hogy egy adott környezetben összesen (az összes vizsgált alaphangfrekvencián együttesen) mekkora volt a nem az ejtési szándék szerinti azonosítások aránya (nazálisok: 39%, zöngés frikatívák: 43%, zöngétlen frikatívák: 44%). A  $\chi^2$ -próba szerint a három hangkörnyezet e paraméter szerint nem egységes ( $\chi^2 = 8,511$ ,  $df = 2$ ,  $p = 0,014$ ), amit a hibák számának (a 18. ábrán is látható) megoszlása alapján úgy értékelhetünk, hogy a nazális hangkörnyezet a zöngés és zöngétlen frikatíva kontextusnál magasabb arányban facilitálta az egyes magánhangzóknak az ejtési szándék szerinti azonosítását. (A nazális környezetben átlagosan 61%, a zöngés frikatívák között átlagosan 57%, míg a zöngétlen frikatívák között 56% volt az ejtési szándék szerinti azonosítás).

A hangkörnyezetek között némiképp különbséget jelent a nem az ejtési szándék szerint megjelölt válaszok között leggyakoribb magánhangzó-minőségek megoszlása. Az 5. táblázatban összeített adatok szerint míg frikatíva környezetben a nem az ejtési szándék szerinti azonosításokkor a zárt /y:/ a leggyakoribb válasz, addig a nazálisok között ejtett magánhangzókra inkább /i:/ válaszok érkeztek.



18. ábra: A második és harmadik nyíltsági fokú /ε e: ø: o:/ hangzókra az egyes alapfrekvenciákon kapott válaszok megoszlása a magánhangzók nyíltsági fokának függvényében

5. táblázat: A nem az ejtési szándék szerint azonosított hangzók esetében kapott magánhangzóválaszok az egyes hangkörnyezetekben

Nazális			Zöngés frikatíva		Zöngétlen frikatíva	
Magánhangzó	Válasz aránya (%)		Magánhangzó	Válasz aránya (%)	Magánhangzó	Válasz aránya (%)
1. /i:/	19,29		/y:/	22,38	/y:/	32,60
2. /o:/	17,91		/o:/	15,49	/o:/	12,90
3. /e:/	17,16		/ø:/	15,15	/ø:/	13,38
4. /y:/	16,14		/u:/	14,80	/i:/	9,26
5. /u:/	12,40		/e:/	12,22	/a:/	7,03
6. /a:/	6,89		/i:/	8,09	/ɒ:/	6,35
7. /ɒ:/	6,50		/ɒ:/	7,06	/u:/	5,49
8. /ø:/	1,99		/a:/	3,10	/e:/	5,32
9. /ε:/	1,77		/ε:/	1,72	/ε:/	0,69

### 2.2.3. Következtetések

Az imént bemutatott kísérletben azt a kérdést vizsgáltuk meg, hogy az előző vizsgálatból származó eredményeink mennyiben írhatók pusztán a vizsgálatban használt nazális mássalhangzó-környezet rovására. Az előző vizsgálat a jelen kísérletben is ellenőrzött főbb eredményei a következők voltak. 1. Az alapfrekvencia emelésével egyre csökkent az egyes magánhangzó-minőségek az ejtési szándéknak megfelelő azonosítása – ezzel párhuzamosan a válaszok egyre nagyobb szóródást mutattak. 2. A vizsgált közepesen magas alapfrekvenciákon a zártabb és nyíltabb szomszédokkal is rendelkező énekelt magánhangzókat a nem az ejtési szándék szerinti azonosítások alkalmával a hallgatók inkább zártabbnak azonosították. 3. A magas alapfrekvencián legnagyobb arányban azonosított magánhangzók nem a legnyíltabb (legnagyobb nyíltsági fokú) magánhangzók voltak, hanem zárt, első nyíltsági fokú magánhangzók (az /i:/ és /y:/), valamint egy harmadik nyíltsági fokú magánhangzó (az /ε/).

Az alfejezetben bemutatott kísérlet kiindulópontját a beszédben megvalósult nazalizált magánhangzók percepciójával kapcsolatos szakirodalom képezte. Eszerint bár a nazális mássalhangzó a koartikulációs nazalizációs hatáson keresztül megváltoztathatja a szomszédos magánhangzó hangszínét, a bejósolható helyen (nazális mássalhangzók környezetében) történt magánhangzó-nazalizációt a percepciós rendszer hatékonyan kompenzálja, így az azonosítás az obstruensek között ejtett nem nazalizált magánhangzókétól nem tér el jelentősen (Krakow et al. 1987). Az éneklés szakirodalmában a nazális hangszín megjelenését tekintve két, egymással ellentétes, vetélkedő nézet létezik: az egyik szerint az éneklésben nincs vagy csak nagyon csekély mértékű a nazális csatolás (Vennard 1964, Yangisawa et al. 1990, Yangisawa et al. 1991), míg a másik szerint énekestől és magánhangzó-minőségtől függően változatos mértékben, de jelen van az orrüregi hatás az éneklésben (Gramming et al. 1993). Bár a nazalitás kérdése az éneklésben vitatott, és az említett vizsgálatok egyike sem ellenőrizte a nazalitás percepciós vetületét. Így Krakow és munkatársainak (1987) percepciós kísérletére alapozva a jelen vizsgálat elején azt feltételeztük, hogy a nazális és a frikatív mássalhangzó-környezetek között nem találunk jelentősebb különbségeket az énekelt magánhangzók percepciójában.

Az eredmények csak részben igazolták az elvárásainkat. Általánosan ugyanis azt láttuk, hogy nem volt jelentősebb eltérés az egyes hangkörnyezetek között, apróbb különbségeket azonban



több ponton is találtunk, és ezek egy része éppen a nazális hangkörnyezet pozitív hatását mutatta az énekelt magánhangzók észlelésében a frikatívákkal szemben.

1. Az adatok szerint az adatközlők az alapfrekvencia emelésével egyre kisebb számban azonosították a magánhangzókat az ejtési szándék szerint kontextustól függetlenül, úgy, hogy az egyre magasabb alapfrekvenciák a válaszok egyre nagyobb szóródását, tehát a magánhangzó-minőségek egyre fokozódó bizonytalanságát mutatták. Ezzel együtt azt láttuk, hogy a nazális hangkörnyezetben ejtett magánhangzók a vizsgált köztes magasabb (500 és 550 Hz) énekelt alapfrekvenciákon valamivel nagyobb arányban idézték elő az ejtési szándék szerinti azonosítást. Az összes azonosítási arányt a három mássalhangzó-környezet szerint összesítve azt találtuk, hogy a nazális hangkörnyezet valamivel magasabb az ejtési szándék szerinti azonosítást mutat. Kiemelendő, hogy a beszédben Krakow és munkatársainak (1987) eredményeihez hasonlóan azt találtuk, hogy a nazális és obstruens környezetben nincs különbség a magánhangzók az ejtési szándék szerinti azonosításának arányaiban.

2. A második és harmadik nyíltsági fokú (zártabb és nyíltabb szomszédokkal egyaránt rendelkező) /ε e: ø: o:/ magánhangzók az éneklésben tapasztalt nem az ejtési szándék szerinti azonosítási tendenciáinak vizsgálatakor azt láttuk, hogy e magánhangzókat a nem az ejtési szándék szerinti azonosításokkor javában zártabbként azonosították a hallgatók, ismét csak kontextustól függetlenül. Emellett egyik kontextusban sem láttunk szoros összefüggést az alapfrekvencia és a nem az ejtési szándék szerinti azonosításokat jellemző tendenciák között. Úgy véljük, a jelen kísérlet bizonyítékkal szolgál arra, hogy a korábbi szakirodalomban eddig nem ismertetett, az észlelést az énekelt magánhangzók esetében jellemző tendencia – mely szerint a zártabb és nyíltabb szomszédokkal is rendelkező magánhangzókat közepesen magas alapfrekvenciákon a nem az ejtési szándék szerinti azonosítások esetében inkább zártabbként azonosítják a hallgatók – a magas alapfrekvencián énekelt magánhangzók azonosításának egy igen fontos jellemzője, ami a beszéd-percepció és a produkció közti nem egyszerű és lineáris összefüggésekről árulkodik. A korábbi percepciós vizsgálatok nagy hangsúlyt fektettek azon magánhangzók elemzésére, melyeket – első nyíltsági fokuk lévén – a hallgatók a nyíltsági fok nem az ejtés szerinti azonosításakor nem is azonosíthattak zártabbként, csak nyíltabbként (Hollien et al 2000), illetve nem adtak pontos, számszerű információkat arról, mire alapozták azt a feltételezésüket, hogy az alapfrekvencia eme-

lését az egyes magánhangzók nyíltabbként azonosítása jellemezné, miközben utaltak rá, hogy egyes zárt, azaz első nyíltsági fokú magánhangzók (elsősorban az /i:/) azonosítása kiemelkedően magas volt még magas alapfrekvenciákon is (Scotto di Carlo–Germain 1985). Ennek megfelelően e tanulmányok, bár megfogalmazták a produkció és percepció egyszerű, lineáris összefüggéseit – mely szerint a produkciós nyitódás egyúttal nyíltabb észleletet is eredményez – valójában nem alapozták meg maradéktalanul e megállapításukat. A 2.1. és 2.2. fejezetben bemutatott vizsgálataink azonban szisztematikus, a magánhangzórendszer sajátosságainak figyelembevételével elvégzett elemzések révén bizonyítékkal szolgáltak arra, hogy a produkció és a percepció közti kapcsolat az énekelt magánhangzók azonosításában sem tekinthető egyszerű, lineáris összefüggésnek (sőt azt az egyszerűbb lineáris összefüggések alapján egyáltalán nem értelmezhető tendenciák jellemzik). Mindez erős bizonyítéknak tűnik a beszédpercepció szakirodalmában már több évtizede ismeretes, a produkció és a percepció bonyolultabb összefüggéseiről szóló ismeretekre is (vö. Srulovicz–Goldstein 1983, Delgutte–Kiang 1984, Chistovich–Lublinskaja 1979, Carlson et al. 1970, Traunmüller 1981 és a jelen értekezés 1.1.2.2. alfejezete).

3. Az előzőekkel is összefüggő módon a jelen vizsgálatban elemzett legmagasabb alapfrekvencián nyíltabb és zártabb minőségű hangzók is voltak azok között, melyeket nagyobb arányban azonosítottak az ejtési szándék szerint a tesztben részt vevők (a nazálisok esetében: /a:/, /u:/ és /e:/, a zöngés frikatívák esetében: /a:/, /u:/ és /y:/, a zöngétlen frikatívák esetében: /y:/ és /o:/). E szempont szerint a fonetikailag zöngés nazális, valamint a nyelvileg is zöngés frikatíva valamelyest közelebb állt egymáshoz, mint a más szempontok szerinti összehasonlításokban.

A vizsgálat kiinduló feltételezése szerint a nazális hatást mint az ejtési szándék szerinti azonosítást akadályozó paramétert teszteltük énekes adatközlőnk produkciójában. Azt a feltételezést, hogy a nazális hangkörnyezet valóban ilyen hatást fejtene ki a magánhangzók azonosítására, eredményeink alapján egyértelműen elutasíthatjuk. Ráadásul a nazális mássalhangzók környezetében ejtett magánhangzók nemcsak, hogy nem mutattak alacsonyabb azonosítási arányokat, hanem sokszor éppen hogy magasabb arányban voltak az ejtési szándék szerint azonosíthatók, mint a frikatívák között ejtett, magas alapfrekvenciákon énekelt magánhangzók. Ennek értelmében tehát megfogalmazhatjuk azt a sejtést is, hogy a nazális mássalhangzók nem pusztán a „zengető” hatás miatt (vö. Kerényi 1959, Váradi 2010) lehetnek kitüntetett szerepűek az éneklésben. Ha

ugyanis a magánhangzók percepciójának elkülönítése a nazális mássalhangzók környezetében bizonyos esetekben (bizonyos alaphangfrekvenciák esetében) jobban biztosított, akkor az énekesek (és énektanárok) számára a nazálisok használata által megerősítést nyerhet az eddigi adataink alapján inkább csak illuzionisztikusnak tűnő, de az énektechnika szempontjából valószínűleg nagyon fontos élmény és érzemény, mely szerint az egyes magánhangzók minősége egymástól elkülöníthető a magas alaphangfrekvenciájú éneklésben. Mivel e feltételezés tesztelésére előbb vizsgálni kellene az énekelt magánhangzók nazalizációjának lehetséges mértékét, majd ennek ismeretében a különböző realisztikus mértékben nazalizált magánhangzók észleletének alakulását is, ezért a felvetés ellenőrzése messze túlmutat a jelen kutatás keretein, és e helyütt csak feltételezés maradhat. Fontos kiemelnünk, hogy a jelen kísérletben részt vevő énekes üregi rezonanciái, illetve a nazalitással kapcsolatos produkciós stratégiája (tehát a nazális csatolás az énekes egyéni fiziológiai és produkciós jellemzői tekintetében) feltehetően sajátos módon, egyénien érvényesültek. Ez azt jelenti, hogy énekes adatközlőnk nazalizációs stratégiái nem feltétlenül képezik le az operai énektechnikában elvárt vagy tradicionálisnak tekinthető stratégiát; miközben azt sem szabad elfelejtenünk, hogy egyes korábbi vizsgálatok sugalmazásai szerint nem is teljesen valószínű, hogy ilyesfajta egységes nazalizációs stratégia feltételezhető az éneklésben (vö. Gramming et al. 1993). A jelen vizsgálatnak tehát nem volt, de nem is lehetett célja a nazalitás megjelenésével kapcsolatos általános érvényű produkciós stratégiákat feltérképezni – kivált, ha elfogadjuk az imént említett tanulmány javaslatát a nagy egyéni variabilitásról. Célunk ezzel szemben csak az lehetett – és ennek a kísérlet, úgy véljük, maradéktalanul eleget is tett –, hogy bemutassuk, hogy az előző kísérletünkben, az ugyanezen énekes adatközlő produkálta magánhangzók azonosításában talált, a produkció és percepció nem egyszerű, lineáris összefüggéseit mutató eredmények egyértelműen nem a nazális hangkörnyezet, hanem a magas alaphangfrekvenciájú éneklés mint produkciós mód sajátosságainak következtében alakultak a tapasztaltak szerint.

A megelőző és a jelen kísérletünk eredményei alapján, úgy véljük, megfogalmazható az a kijelentés, hogy nincs okunk feltételezni a produkció és percepció közti (korábbi vizsgálatokban sugallt) nagyon leegyszerűsítő összefüggést, mely szerint az éneklés során az alaphangfrekvencia emelésével egyre fokozódó produkciós nyitódást (azaz az állkapocsnyitás szögének növelését, illetve az F1 emelését) egyértelműen percepciósnak nyitódás is kísérné. Az adatok alapján sokkal inkább azt

mondhatjuk, hogy az alapfrekvencia emelését kísérő produkciós változások hatására a magánhangzók percepció minősége egyre bizonytalanabb lesz az éneklésben (eddig adataink alapján a 650 Hz közepesen magas alapfrekvenciáig). Az egyes hangzóminőségek az alapfrekvencia emelésével egyre fokozódó percepció változásairól ezen felül úgy láttuk, hogy azok graduális változást jelentenek – az alapfrekvencia emelésével egyre jobban megváltozó akusztikai szerkezetű magánhangzók tehát nem kategorikusan az egyik magánhangzó-kategóriáról a másikra ugrásként csapódnak le az észleletben.

A következő kísérletünkben az éneklés szakirodalmában azt a gyakran empirikus bizonyítékok nélkül is elfogadott feltételezést kívánjuk megvizsgálni, mely szerint a koartikuláció akusztikai vetületének, azaz a formánsátmeneteknek köszönhetően a magas alapfrekvencián énekelt magánhangzók a beszédhangokhoz hasonlóan magasabb arányban azonosíthatók az ejtési szándék szerint, mint az izoláltan ejtett magánhangzók. Mivel eddigi akusztikai eredményeink azt mutatták, hogy az énekelt magánhangzók akusztikai szerkezete megváltozik az alapfrekvencia emelésével, percepció eredményeink szerint pedig ez a változás elbizonytalanítja a magánhangzók észlelt minőségét, úgy véljük, hogy a vizsgálandó feltételezés bonyolultabb annál, mintsem, hogy a beszédpercepció szakirodalmából ellenőrzés, azaz szisztematikus és jól kontrollált kísérletekből nyert empirikus bizonyítékok nélkül adaptálható legyen.

### **2.3. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: azonosítás mássalhangzó-környezetben, izolált ejtésben, valamint a zöngéindítás mint akusztikai kulcs hiánya esetén**

Mint azt a korábbi, valamint a jelen értekezés *2.1. alfejezetében* bemutatott produkciós vizsgálatokban láttuk (továbbá amint azt a *2.5. alfejezetben* bemutatandó kísérletünk is demonstrálni fogja), az énekelt magánhangzók produkciója igen erősen függ az alapfrekvenciától, olyan módon, hogy a magánhangzók artikulációs és akusztikai megvalósítása az alapfrekvencia emelésével egyre inkább eltér a magánhangzók beszédbeli megvalósításától. Amint azt az eddig bemutatott szakirodalmi leírások és saját észlelési kísérleteink bizonyítják, e produkciós változás – termé-

szetszerűleg – a magánhangzók észleleti vetületét is erősen befolyásolja, általánosságban olyan módon, hogy az alaphfrekvencia emelésével egyre csökken az ejtési szándék szerinti magánhangzó-azonosítások aránya. Mégis, mindennapi tapasztalataink azt mutatják, hogy a nem laboratóriumi körülmények között rögzített, nem mesterségesen generált, értelmetlen hangsorokból álló stimulusok, hanem az adott nyelven értelmes szavakká, majd értelmes szöveggé formálódó beszédhangok éneklése során jó arányban lehetséges a magánhangzóknak az ejtési szándék szerinti azonosítása (hiszen többé-kevésbé megértjük a dalok és operaáriák szövegét). Jogosan merül fel tehát a kérdés, hogy mennyiben segítheti a mássalhangzó-környezet és nagyobb kontextusban az értelmes szavakba és mondatokba ágyazottság az egyes magánhangzók azonosítását.

A kontextus szerepe az énekelt magánhangzók észlelésében, ahogy a beszédpercepcióban is, valójában kétarcú probléma. Az énekelt magánhangzók észlelését eddig vizsgáló szakirodalomban azonban az látható, hogy e két, élesen elkülönítendő oldal nem válik szét hangsúlyosan (erről már a szakirodalmat bemutató *1.3.3.2. fejezetben* is értekeztünk részletesebben, de a későbbiekben még több ponton utalunk rá). A kontextus kérdésének egyik aspektusa a pusztán a beszédmegértés legalsóbb szintjeire támaszkodó bottom-up folyamatok aktivációjával kapcsolatos beszédészlelési probléma: miként és mennyiben segíti a magánhangzó-azonosítást a magánhangzóknak a szomszédos mássalhangzókkal mind időben, mind térben átfedő, együttes ejtésének, azaz a koartikulációnak az akusztikai vetülete, a koartikulációs formánsátmenet? A kérdés másik aspektusa azonban nemcsak az akusztikai elemzést végző bottom-up folyamatokra támaszkodó, de a beszédmegértési folyamat felsőbb szintjeinek (tehát a top-down folyamatok) aktivációjához is köthető észlelési probléma. Eszerint ha a magánhangzó-felismerés értelmes, valamint esetleg szintaktikailag és szemantikailag is beágyazott szavakban zajlik, a lehetséges – tehát a felismerés kimeneteként szóhajóhető – magánhangzóválaszok száma és minősége jóval korlátozottabb, mint a nem értelmes, izolált szavak (vagy izolált hangzók) esetén. Ennek megfelelően ezekben a felismerési helyzetekben valójában egyáltalán nem szükséges a maradéktalan akusztikai dekódolás az egyes magánhangzók minőségének felismeréséhez – az akusztikailag nem beazonosítható magánhangzókat képesek vagyunk „kitalálni” a részben vagy egészben dekódolt kontextus alapján.

A mássalhangzó-környezet mint a magánhangzóknak az ejtési szándék szerinti azonosítását segíthető jelenség tehát két, egymástól eltérő aspektus mentén is vizsgálendő jelenség. A jelen

alfejezetben bemutatott vizsgálat e két aspektus közül az elsőre fókuszál, a kísérlet fő célja annak a kérdésnek a megválaszolása, hogy mennyiben segíti a mássalhangzó-környezet a formánsátmenetek révén a magas alapfrekvencián énekelt magánhangzók az ejtési szándék szerinti azonosítását. A kérdés másik aspektusát, egészen pontosan a két aspektus közti különbségek módszeres vizsgálatát, azaz a formánsátmenetek és az értelmes szavakba ágyazottság hatásának összevetését a következő alfejezetben ismertetett kísérletünkben teszteljük. A jelen kísérletben ugyanakkor még egy további kérdésre is választ keresünk, ugyanis az énekelt magánhangzók esetében is meg kívánjuk vizsgálni azt a pszichoakusztikai szakirodalomban leírt jelenséget, mely szerint a hangszín azonosításában kiemelt szerepe van a hangindításnak.

Gottfried és Chew (1986) kísérletében mássalhangzó-környezetben ejtett magánhangzók azonosítását vetette össze a kontextusból kivágott magánhangzó-tisztafázisok észlelésével. Bár éppen emiatt a kísérlet eredményei nem vonatkoztathatók az izoláltan ejtett magánhangzókra és azok, valamint a mássalhangzó-környezetben ejtett magánhangzók észlelési különbségeire, a kísérleti elrendezés mégis felhívja a figyelmet egy izgalmas és tesztelendő kérdésre. Egyes pszichoakusztikai kísérletek eredményei szerint ugyanis egy adott alapfrekvencián megszólaló hang hangszínének azonosításában, azaz a hangszerek keltette hangok világában a hangot kiadó hangszer felismerésében kiemelt szerepű akusztikai kulcs a hangindítás, ezen információ hiányában ugyanis jelentős mértékben csökken a hangszerek felismerésének sikeressége (Berger 1964, Saldanha–Corso 1964). Felmerülhet tehát a kérdés, hogy vajon a hangadás kezdetének kiemelt jelentősége van-e a beszédhangok világában is, ahol az egyes hangszínbeli különbségeket az egyes magánhangzó-minőségek különbségeinek feleltethetjük meg. Bár Gottfried és Chew (1986) a magánhangzók tiszta fázisának tesztelésével gyakorlatilag megvizsgálja a kérdés egyik oldalát (ez volna a hangindítás eliminálásával kapott stimulusanyag), és összeveti azt a kontextusban ejtett magánhangzók azonosításával, nyitott kérdés marad, hogy milyen eredményekre vezetne a magánhangzók tiszta fázisára kapott észlelési adatok összevetése az izolációban ejtett énekelt hangzók azonosításával. E kérdés vizsgálata képezi tehát a jelen alfejezetben bemutatásra kerülő vizsgálat második célját.

Mielőtt rátérnénk a kísérlet anyagának és módszertanának ismertetésére, fontosnak tartjuk emlékeztetőül röviden ismét áttekinteni az *1.2.2. alfejezetben* ismertetett, a szomszédos mássalhang-

zók hatását az énekelt magánhangzók azonosításában vizsgáló szakirodalom problematikus pontjait, melyek alapján úgy véljük, hogy a mássalhangzó-környezet hatása az énekelt magánhangzók észlelésében nem tekinthető sem alaposan körüljárt, sem pedig megnyugtatóan megválaszolt kérdésnek.

Azon feltételezés tudományos igazolására, hogy a mássalhangzó-környezet megléte vagy hiánya hatással van-e a magas alapfrekvencián (szoprán éneklésben) ejtett vokálisok azonosítására, eddig összesen egyetlen kutatás született. (Egy további vizsgálat, a már említett Gottfried és Chew 1986 szerzőpáros munkája mélyebb alaphangokon, egy férfi hangkategóriában, egy tenor énekes ejtésében vizsgálta a kérdést.) Smith és Scott (1980) a magas alaphangon énekelt magánhangzók azonosítását *bVd* hordozóhangsorban vizsgálta egy szoprán énekes ejtésében. Percepció adataik a szerzők értelmezése szerint az operanéklésre nézve is megerősítik azt a beszédben tapasztalható tendenciát, miszerint a szomszédos mássalhangzók segítik a magánhangzók felismerését. Adataikat a szerzők azonban nem szigorúan kontrollált körülmények között nyerték, a módszertan több okból is nehezen értelmezhető eredményekhez vezetett.

Egyrészt a magánhangzót közrefogó mássalhangzók képzéshelye eltérő volt, így valójában nem lehet biztosan tudni, hogy mely mássalhangzó befolyását tesztelték a szerzők, sem pedig azt, hogy van-e különbség a magánhangzót megelőző és követő hangátmenetek hatása között.

Másrészt az sem egyértelmű, hogy milyen módon kontrollálták a szerzők az énekes gégejének függőleges helyzetét, mely állításuk szerint szintén fontos befolyásoló tényezőnek bizonyult. A kísérlet leírásából úgy sejthető, a „gégehelyzet” pusztán csak egy, a szerzők egyéni percepciók megítélésén alapuló paraméter, tehát valamilyen percepció alapú, de nem pontosan definiált kategória volt.

Egy harmadik probléma, hogy a szerzők feltételezése szerint az  $F1 : f0$  formánshangolás mint az operai énekmód sajátossága módosítja a beszédben jellemző akusztikai szerkezetet, és így csökkenti az ejtési szándék szerinti azonosíthatóságot; ezzel magyarázható, hogy az izolált ejtésű magánhangzók az ejtési szándék szerinti azonosításának aránya alacsonyabb volt az operai módban a „nem rendes” énekprodukciónál képest. Ha viszont ennek ellenére a mássalhangzó-környezetben mégis magasabb a magánhangzók ejtés szerinti azonosítása, mint izolált ejtésben, akkor az azt jelenti, hogy a szerzők szerint valójában a formánsstratégia, azaz az  $F1 : f0$  hangolás

érvényesült a kontextus (tehát a mássalhangzó-környezet megléte vagy hiánya) függvényében, ami miatt a két kondícióban megvalósult magánhangzók akusztikai szerkezete eltérő lehetett, és a hallgatók különböző minőségű magánhangzóként ismerhették fel őket.

Egy negyedik, és a vizsgálat eredményeinek értelmezését legjobban megnehezítő probléma pedig a kísérlet alapvető felépítésének, módszertanának egy meghatározó sajátossága. Míg a teszt során a résztvevők a CVC kondícióban négy lehetőségből, négy értelmes szó közül választhattak, addig az izolált ejtésű hangok esetében a választási lehetőségek egyáltalán nem korlátozódtak arra a négy hangzóra, amit az értelmes szavak megengedtek. Az izolált ejtésben ugyanis a válaszlehetőségek számát csak az angol nyelv magánhangzókészlete korlátozta, mely legalább tíz lehetséges magánhangzót választ hagyott. Ennek fényében tehát Smith és Scott (1980) a magánhangzóknak a mássalhangzó-környezet meglététől függő azonosítását valójában két jelentősen eltérő percepciós működést összevetve, két jelentősen eltérő kondícióban igyekezett vizsgálni – éppen a fent említett kettősség értelmében. Eszerint az egyik (CVC) kondícióban az értelmes szavak révén a nyelv magasabb szintjei, a morfológia, illetve a szemantika is befolyásolta az azonosítás tendenciáit, hiszen ezek korlátozták a válaszlehetőségeket és a válaszlehetőségek számát a válaszként szóba jöhető létező szavak (azaz a top-down percepciós folyamatok) révén. Ám a másik, az izolált ejtésű hangzókat tartalmazó kondícióban ezzel szemben csak az alacsonyabb szintű, elemi hallási észlelésen és az akusztikai elemzésen alapuló (és csak az akusztikai kulcsokra támaszkodó) beszédpercepciós (bottom-up) folyamatok léphettek működésbe úgy, hogy a lehetséges válaszok számát az angol nyelv magánhangzókészletén kívül semmi sem korlátozta. Vizsgálatukban a szerzők tehát nem választották szét a mássalhangzó-környezet két szigorúan elkülönítendő aspektusát, és így bár eredményeik a beszédmegértés magasabb folyamatainak aktivációjából is származó percepciós működésekről tanúskodtak, Smith és Scott a tapasztalt pozitív hatásokat kizárólag a formánsátmenetek mint akusztikus kulcsok javára írta.

Mindezek értelmében Smith és Scott (1980) kísérlete valójában nem ad egyértelmű választ a mássalhangzó-környezet hatásának kérdéseire, ráadásul a vizsgálati elrendezés a kondíciók több ponton eltérő feltételei miatt a vizsgálati eredmények több, egymással interakcióban lévő zavaró (confound) hatás eredményeként álltak elő (melyek támogatták, erősítették a különbségeket a



mássalhangzó-környezet hatását reprezentálni kívánó kondíció és az izolált ejtést reprezentáló kondíció között).

A jelen alfejezetben ismertetett kísérlet feltételezése (az *1.3. alfejezetben* is megfogalmazott hipotézissel összhangban) a következő. Abból kiindulva, hogy az  $f_0$  emelésével a spektrális alul-mintavételezettség fokozódik, illetve az  $F1 : f_0$  hangolás következtében egyre inkább megváltozik az énekelt magánhangzók akusztikai szerkezete az alapfrekvencia emelésével, ezért a magas alapfrekvenciákon énekelt magánhangzók azonosítása egyre kevésbé történhet az ejtési szándék szerint. Mivel azonban az alapfrekvencia emelésével egyre fokozódó változás magukat a magánhangzó-megvalósulásokat (tehát a produkciót), nem pedig a percepciót érintő folyamat, feltételezzük, hogy sem a mássalhangzó-környezetben rejlő dinamikus információ, sem pedig a zöngéindítás nem segítheti a percepció működéseket abban, hogy ezeket a folyamatokat kompenzálják, és a magas alapfrekvencián énekelt magánhangzókat az ejtési szándék szerint azonosítsák a megvalósított, az  $f_0$  emelésével egyre bizonytalanabb minőségű magánhangzók helyett. Az alacsonyabb alapfrekvenciák (azaz a beszédbeli megvalósuláshoz hasonlóbb akusztikai minőségű énekelt magánhangzók) esetében ugyanakkor azt feltételezzük, hogy a mássalhangzó-környezet bizonyos mértékű pozitív hatása lesz kimutatható a közrezárt magánhangzónak az ejtési szándék szerinti azonosításában. A zöngéindítás ennél valamelyest enyhébb, de mégis pozitív hatását feltételezzük, mely (az alacsonyabb alapfrekvenciákon) a zöngéindítás nélkül prezentált hangok azonosításához képest valamivel magasabb az ejtési szándék szerinti azonosítási arányokat eredményez.

### **2.3.1. Anyag, módszer és kísérleti személyek**

A vizsgálathoz egy professzionális szoprán énekesnő (életkora 27 év) énekprodukciónak rögzítettük. Az énekes a magyar nyelv három, akusztikai tekintetben egymástól legtávolabb eső magánhangzóját /a: i: u:/ ejtette izoláltan („V” kondíció) és  $bVb$  hangkörnyezetben („CVC” kondíció), hat énekelt alaphangmagasságon (ezeket a 6. táblázatban összesítettük) és beszélve (melynek átlagos alaphangmagassága 191 Hz volt). A vizsgálat újítása eddigi produkciós és percepció kísérleteinkhez képest, hogy ez alkalommal már a temperált zenei skálák figyelembevételével választottuk meg a vizsgált alapfrekvenciákat, olyan módon, hogy azok (hangmagasság-érzetük

szerint) egymástól egyenlő távolságokra helyezkedjenek el, valamint gyakorlatilag teljesen lefedjék a szoprán énekesek által használt alapfrekvencia-tartományt.

6. táblázat: A kísérletben használt zenei hangok és azok frekvenciaértéke

Zenei hang	Alapfrekvencia (Hz)
$f$	175
$h$	247
$f'$	349
$h'$	494
$f''$	698
$h''$	988

A teszthangsorokon kívül disztraktor hangsorokat is rögzítettünk ugyanazon énekessel. A disztraktorok további magyar magánhangzókat /ɒ ɛ e: o: ø: y:/ tartalmaztak *bVb* hangkörnyezetben és izolált ejtésben, valamint beszélve, illetve a  $f'$ ,  $h'$ ,  $f''$  zenei alaphangokon énekelve. Azonban ezen stimulusok percepcióis eredményeit nem elemeztük.

A hordozó hangsor megválasztását a következők motiválták. A mássalhangzó-környezet szigorú kontrollálása érdekében olyan CVC hordozóhangsort használtunk, melyben a két mássalhangzó azonos, így az adataink egyértelműen az adott képzéshelyű mássalhangzóra vonatkoztathatóak. Azonban ahhoz, hogy az eredményeinket a Smith és Scott (1980) kísérletében látottakkal is összehasonlíthatóvá tegyük, az általuk használt *bVb* hangsor egyik mássalhangzóját kellett megtartanunk. Korábbi kutatási tapasztalataink szerint a CVC hangsor első mássalhangzója gondosabban artikulált az éneklésben, mint az utolsó, ezért a *bVb* hangsor első mássalhangzójának a megtartása mellett döntöttünk.

A magánhangzó zöngéindulásának hatását egy harmadik kondíció előállításával teszteltük: egy hangedítáló program, a WaveSurfer (Sjölander–Beskow 2000) segítségével („fokozatos beúszás” hatás alkalmazásával) eltávolítottuk az izoláltan ejtett hangok elejét („Vágott” kondíció). A beúsztatást 30 ms-os időtartamon, logaritmikus függvény segítségével valósítottuk meg. A felvételeket csendesített szobában rögzítettük 44,1 kHz-es mintavételezéssel, 16 biten.

A hanganyagokat egy, a Praat programban (Boersma–Weenink 2009) írt és futtatott percepcióis tesztben vizsgáltuk 22 ép hallású felnőtt részvételével (21 és 35 év közöttiek, átlagos életkoruk 29 év). A teszt (3 kondíció × 3 magánhangzó × 7 alaphangmagasság =) 63 stimulust és 30

disztraktort tartalmazott. Minden tesztelő minden stimulust kétszer hallott, véletlenszerű sorrendben, így egy ülésben összesen 156 hangminta hangzott el. A teszt előtt a kísérleti személyek azt az utasítást kapták, hogy azonosítsák az önmagában vagy hangsorban elhangzó magánhangzót, és döntésüket a képernyőn megjelenő kilenc magánhangzó egyikének megjelölésével rögzítsék. Ehhez a képernyőn válaszként a magyar hosszú (vagy hosszan ejthető) magánhangzókat /ɒ a: ɛ e: i: o: ø: u: y:/ jelenítettük meg ortografikus alakban. A tesztelők a stimulusokat binaurálisan, fejhallgatók keresztül hallották, és a számítógéphez csatlakoztatott egérrel kattinthatnak a kiválasztott magánhangzóra. A teszt átlagosan 10-15 percet vett igénybe, mivel azonban a stimulusok sorrendje minden tesztben eltérő volt, nem volt szükséges figyelembe vennünk a fáradás következtében esetlegesen fellépő teszthatást.

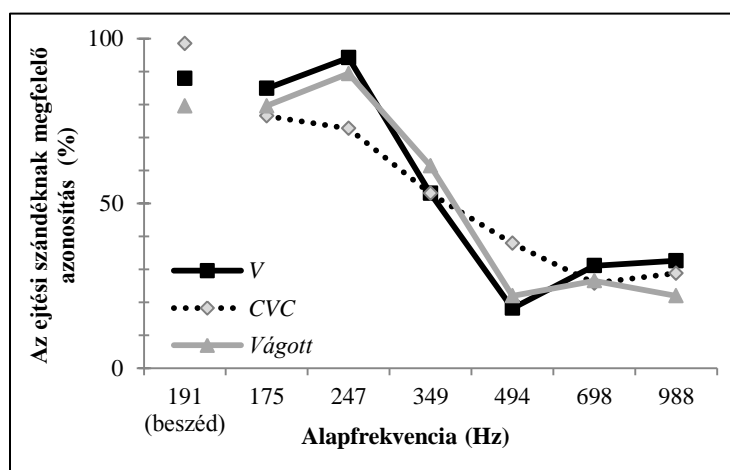
A válaszokat ismét tévesztési mátrixokban összesítettük minden stimulus-válasz párt alapfrekvenciánként rendezve.

Az adatok statisztikai elemzését az R programmal végeztük (R Core Team 2013). Az elemzés során előbb Shapiro–Wilk-próbával elemeztük a normális eloszlás kritériumát, melynek teljesülése esetén parametrikus tesztek alkalmaztak (ANOVA), ellenkező esetben pedig nemparametrikus tesztek ( $\chi^2$ -próba, Mann–Whitney U-próba). A korrelációs számításokat a Pearson-féle korrelációelemzéssel végeztük.

### 2.3.2. Eredmények

A 19. ábra a három vizsgált magánhangzónak az ejtési szándék szerinti azonosításának százalékos arányát ( $y$  tengely) mutatja az alaphangmagasság függvényében ( $x$  tengely), a három kondíció szerint szétválasztva, de mindhárom magánhangzó-minőséget összesítve. Az adatok megjelenítése azt sugallja, hogy a vokálisoknak az ejtés szerinti azonosítása csökken az  $f_0$  emelésével. A kéttényezős ANOVA szerint, melyben az  $f_0$  és a *kondíció* hatását teszteltük, ez a benyomás helyes, és az  $f_0$  paraméter valóban hatással van az ejtési szándék szerinti azonosítási arányok alakulására ( $F(1) = 33,33$ ,  $p < 0,001$ ). Az összes lehetséges összes pár összehasonlítása (Bonferroni-korrekcióval) azt is felfedte, hogy az egymást alaphangmagasság szerint követő énekelt stimulusok között egészen a  $h'$  zenei hangig (494 Hz) mutatkozik páronként szignifikáns eltérés ( $p < 0,02$ ), mely alaphangtól felfelé azonban a különbségek már nem ilyen jelentősek. A Pearson-

féle korrelációs tesztek azt mutatták, hogy az *f0* és az *ejtési szándéknak megfelelő azonosítás arányának* kapcsolatát viszonylag erős negatív korreláció jellemzi. Amennyiben a három magánhangzóra kapott adatokat egy csoportban vizsgáljuk, erős összefüggést találunk:  $r = -0,631$ ,  $p < 0,001$ . Ha pedig szétválasztjuk az adatokat a három kondíció mentén, némileg eltérő erősségű, de minden esetben szignifikáns összefüggéseket látunk: „V”:  $r = -0,590$ ,  $p = 0,005$ ; „CVC”:  $r = -0,609$ ,  $p = 0,003$ ; „Vágott”:  $r = -0,703$ ,  $p < 0,001$ .



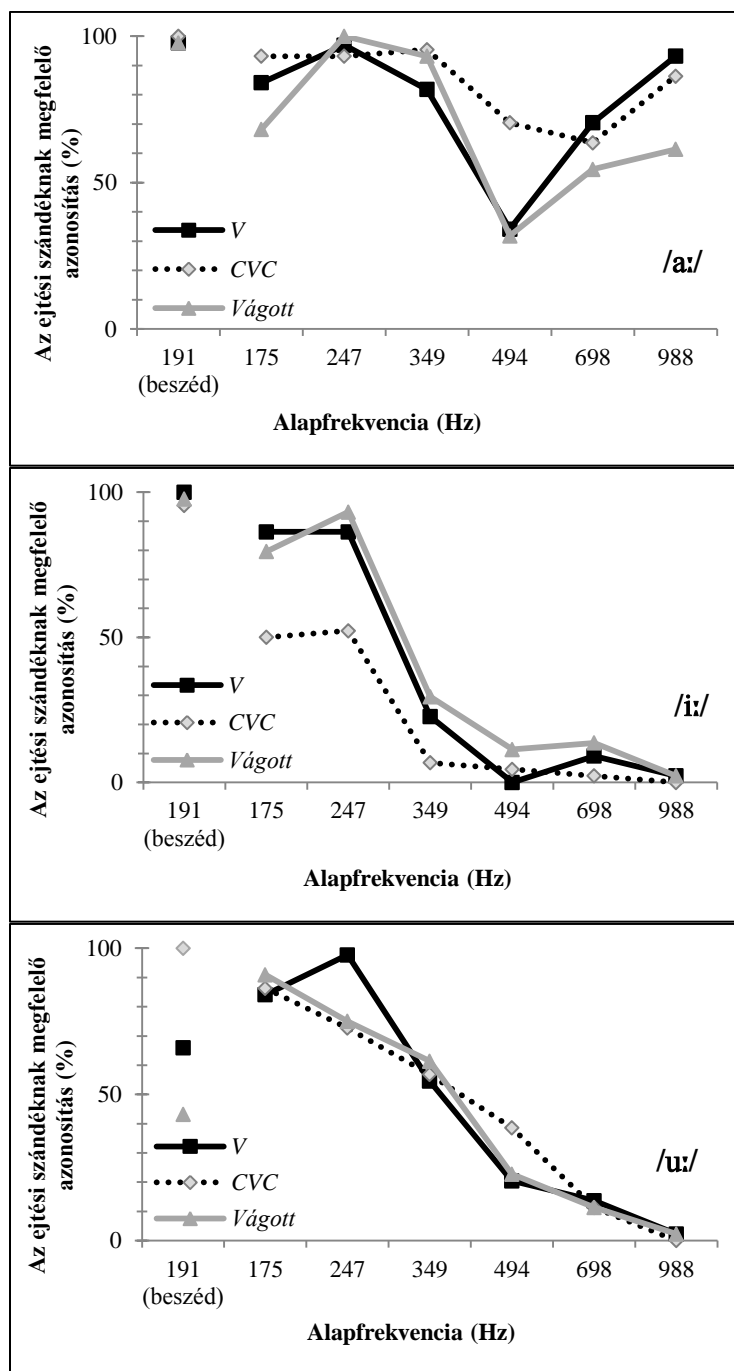
19. ábra: A három vizsgált magánhangzó (/a: i: u:/) azonosításának összesített százalékos aránya az alaphang függvényében

A kéttényezős varianciaanalízis (ANOVA) nem igazolta, hogy a *kondíció* is hatással lenne az észlelési eredményekre (sem önmagában, sem pedig az *f0* változóval kombinálva). Ráadásul a 19. ábrát szemlélve az is kitűnik, hogy a *kondíció* és az *ejtési szándék szerinti azonosítás* közti összefüggés valójában nem is volt konzisztens. Eszerint nemcsak, hogy nem teljesül, hogy a magánhangzókat nagyobb arányban volna lehetséges az ejtési szándék szerint azonosítani mássalhangzó-környezetben, de az is igaz, hogy néha (például a *h* zenei alaphangon) kifejezetten rosszabb a CVC szekvenciában elhangzó vokálisoknak az ejtési szándék szerinti felismerése, mint az izolációban ejtett hangzóké. A különbségek statisztikailag ugyan nem igazolhatók, általánosságban azonban mégis úgy látszik, nagyobb eltérés jelentkezett a „V” és „CVC”, illetve a „Vágott” és a „CVC” minták felismerési aránya között, mint a „V” és „Vágott” minták között.

Bár az éneklésben nem igazolódott a fonetikai kontextus (a mássalhangzó-környezet és a zöngéindulás) hatása, a beszédben az elvárt tendenciákat találtuk: a legmagasabb az ejtési szándék szerinti azonosítási arányt a „CVC” kondíció mutatta, melyet a „V” kondíció (izolált ejtés), és végül pedig a „Vágott” kondíció (eliminált zöngékezdéssel előállított stimulusok) követték.

Ugyan a fonetikai kontextus hatása a magánhangzó-azonosításra nem igazolódott, az  $f_0$ , a *kondíció* és a *magánhangzó-minőség* változókkal lefuttatott varianciaanalízis szerint nemcsak az  $f_0$ , de a magánhangzó minősége ( $F(2) = 22,673$ ;  $p < 0,001$ ), illetve e két faktor kombinációja is ( $F(2) = 7,419$ ;  $p = 0,002$ ) befolyásolta a hangzóazonosítási értékeket. Ezért az ejtési szándék szerinti azonosítási arányok változását a magánhangzó-minőségek mentén is elemeztük.

A magánhangzó-minőségek közti összes lehetséges páros szembenállás összevetése szerint Bonferroni-korrekciónak után is szignifikáns különbség mutatkozott az /i:/ – /a:/ és /u:/ – /a:/ párok, tehát az első és negyedik nyíltsági fokú magánhangzók között ( $p < 0,018$ ), de nem volt különbség a nyíltsági fok tekintetében egyező, első nyíltsági fokú (zárt, felső nyelvállású) /i:/ és /u:/ között. Ezt az elkülönülést rajzolja ki az  $f_0$  és az *ejtési szándék szerinti azonosítási arány* közti korreláció is, ha azt a magánhangzók csoportjain belül vizsgáljuk. A két változó ugyanis csak az /i:/ és /u:/ esetében mutat erős, szignifikáns összefüggést (/i:/:  $r = -0,781$   $p < 0,001$ ; /u:/:  $r = -0,900$ ;  $p < 0,001$ ), míg az /a:/ esetében csak gyenge nem szignifikáns összefüggés látszik ( $r = -0,384$ ;  $p = 0,086$ ). A statisztikai elemzések, továbbá a 20. ábra tanúsága szerint tehát a zárt képzésű vokálisok az ejtési szándék szerinti azonosítása csökkent az alaphang emelésével, míg a nyílt képzésű (negyedik nyíltsági fokú) beszédhanghoz tartozó százalékok az  $f_0$  értékétől megközelítőleg függetlenül alakultak.



20. ábra: Az /a:/, /i:/ és /u:/ az ejtési szándék szerinti azonosításának százalékos aránya az alapfrekvencia függvényében

Az /a:/ az ejtési szándék szerinti azonosítási aránya a  $h'$  zenei alaphang (494 Hz) alatt viszonylag magas volt minden kondícióban, a  $h'$  alaphangnál azonban hirtelen esést tapasztaltunk

(20. ábra). Ezen az alapfrekvencián 46%-ban azonosították a hangzókat az ejtési szándéktól eltérően a tesztelők, és minden esetben a magasabb első formánssal ejtett (zártabb képzésű) /ɒ/-t jelölték meg válaszként. A  $h'$  fölött újabb emelkedés következett az ejtés szerinti azonosítás arányában, majd ezt követően folyamatos emelkedést tapasztaltunk a legmagasabb alaphangig, ahol a „V” kondícióban (izolált ejtés) 93%, a „CVC” kondícióban (mássalhangzó-környezet) 86%, míg a „Vágott” kondícióban (eltávolított zöngéindítás) 61% volt az ejtési szándéknak megfelelő válaszok százalékos aránya. Az  $f''$  zenei alaphang esetében ismét az /ɒ/ hang dominálta az ejtési szándéktól eltérő azonosításokat (az összes válasz 27%-ában), míg a legmagasabb,  $h''$  zenei alaphang esetében az elől képzett /ɛ/ (az összes válasz 8%-ában). Mindkét hangzó zártabb képzésű, és a beszédben magasabb első formánssal realizálódik, mint az ejteni szándékozott magánhangzó.

Az /i:/ az ejtési szándék szerinti azonosítása már az egészen alacsony alapfrekvenciájú  $f'$  hangnál hirtelen csökkenést mutatott (20. ábra), itt és a magasabb alaphangok esetében is az ejtési szándéktól eltérő azonosítások legnagyobb részében /e/-t jelöltek a hallgatók ( $f'$ : 80%,  $h'$ : 94%,  $f''$ : 67% az összes válaszból). A  $h''$ , azaz a legmagasabb zenei hang esetében azonban az /i:/-t is leggyakrabban /a:/-ként ismerték fel a hallgatók (az összes válasz 80%-ában), az ejtési szándék szerinti azonosítások aránya 2% körül alakult (minden kondícióban).

Az /u:/ azonosítását leíró tendenciák a 20. ábrán bemutatott átlageredményekhez hasonlóan alakulnak: az  $f_0$  emelésével fokozatosan csökkenő százalékarányokat láttunk, míg a „V” és „CVC” kondíciók között nem következetesen alakuló kapcsolatok mutatkoznak. Az ejtési szándéktól eltérő azonosításokat tekintve a legtöbb alapfrekvencián leggyakrabban előforduló hangzó az /o:/ ( $f$  és  $h$ : 13%,  $f'$ : 42%,  $h'$ : 62%,  $f''$ : 25% az összes válaszból). Ugyanakkor a legmagasabb alapfrekvencián ( $h''$ ) itt is az /a:/ uralta az ejtési szándéktól eltérő azonosítások hierarchiáját (az összes válasz 86%-ában), és az /u:/ azonosítása csak maximum 2%-ban bizonyult az ejtési szándékkal megegyezőnek.

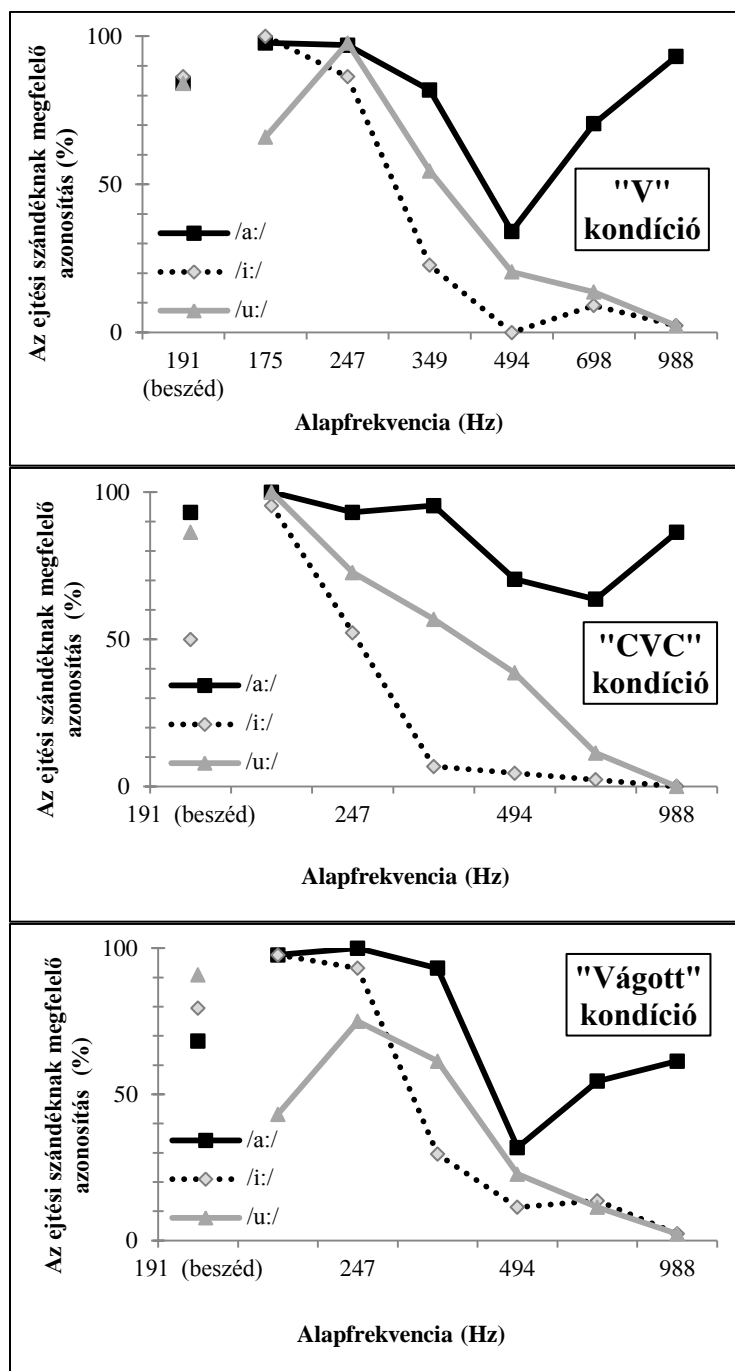
Az /u:/ beszélt ejtésben bemutatott realizációi (az /i:/-vel és /a:/-val szemben) a hipotézisben megfogalmazottaknak megfelelő tendenciát mutatták: a legnagyobb arányban a „CVC” kondícióbeli megvalósulást azonosították a hallgatók az ejtési szándék szerint, melyet a „V” és végül a „Vágott” kondíciók követtek. Azonban ez az /u:/ esetében is csak a beszélt ejtésre korlátozódó

mintázat volt (a többi hang esetében pedig sem a beszédben, sem másutt nem tapasztaltuk ezt a tendenciát).

Összehasonlítottuk az összes lehetséges hangzószembenállást a három kondíció alkotta csoporton belül is. Ezek az összehasonlítások megmutatták az egyetlen olyan tendenciát, mely a jelen adatokban a kondícióval, és azon belül is a mássalhangzó-környezet meglétével is összefüggésbe hozhatónak bizonyult. Egyik csoporton belül sem találtuk ugyanis eltérőnek a különböző nyíltsági fokú magánhangzókat, kivéve a „CVC” csoportot. Itt az eltérő nyíltsági fokú magánhangzók azonosítási arányát eltérő mértékben látszott befolyásolni az  $f_0$  értéke, bár csak a nyelv vízszintes mozgásának tekintetében jobban eltérő pár, az /i:/ és az /a:/ esetében találtunk a Bonferroni-korrektúra után is szignifikáns ( $p < 0,02$ ) eltérést (a nyelv vízszintes mozgásának tekintetében egymáshoz hasonlóbb /u:/ és /a:/ nem tért el). Az összehasonlításokat a 21. ábra grafikonjain szemléltetjük.

Az adatokból jól látható, hogy a különbségek megjelenése a *CVC* kondícióban (illetve meg nem jelenése a többi kondícióban) annak köszönhető, hogy a *CVC* kondícióban tapasztalható leginkább az /a:/ percepciók stabilitása, ugyanakkor a zárt /i:/ is itt mutatja a legalacsonyabb az ejtési szándék szerinti azonosítási arányokat. Míg a *V* és *Vágott* kondíciókban csak a  $h'$  zenei hang fölött, addig a *CVC* kondíciókban már a jóval alacsonyabb  $f$  hang fölött is divergáló tendenciákat látunk a három magánhangzóra.





21. ábra: A magánhangzók az ejtési szándék szerinti azonosításának aránya a három kondícióban

### 2.3.3. Következtetések

A jelen alfejezetben bemutatott vizsgálat célja kettős volt, egyfelől a mássalhangzó-környezet, másfelől pedig a zöngéindítás mint az ejteni szándékozott magánhangzó-minőség felismerését segítő akusztikus kulcs tesztelését céloztuk meg. Amint azt a 2.3. *alfejezet* bevezetőjében is hangsúlyoztuk, a mássalhangzó-környezet hatásának kérdése valójában kettős, és a jelen vizsgálat a két aspektus közül csak az elsőre fókuszált. Első kérdésünk ennek értelmében az volt, hogy valóban segíti-e, és ha igen, milyen mértékben a mássalhangzó-környezet a koartikulációs formánsátmenetek révén a magas alapfrekvencián énekelt magánhangzónak az ejtési szándék szerinti azonosítását egy, a csak a bottom-up akusztikai elemzésre támaszkodó percepciós feladatban. A vizsgálat második kérdéseként korábbi pszichoakusztikai kísérletek eredményeire alapozva azt fogalmaztuk meg, hogy vajon az izoláltan ejtett magánhangzók zöngéindulási szakasza is hozzásegítheti-e a hallgatókat az ejteni szándékozott magánhangzók minőségének azonosításához. A vizsgálat elején azt feltételeztük, hogy mivel az alapfrekvencia emelésével egyre növekszik a spektrális alul-mintavételezettség mértéke, valamint azért, mert a magas alapfrekvenciákon énekelt magánhangzók esetében  $F1 : f_0$  hangolást feltételezünk, magas alapfrekvencián sem a formánsátmenetek, sem pedig a zöngéindulás nem segítheti az ejtési szándék szerinti magánhangzók azonosítását, hiszen az éneklésben magas alapfrekvencián az ejtési szándéktól eltérő akusztikai szerkezetű, illetve az  $f_0$  emelésével egyre bizonytalanabb minőségű hangzók valósulnak meg. Az alacsonyabb alapfrekvenciák (azaz a beszédbeli megvalósuláshoz hasonlóbb akusztikai minőségű énekelt magánhangzók) esetében ugyanakkor azt is feltételeztük, hogy a mássalhangzó-környezet bizonyos mértékű pozitív hatása kimutatható lesz a közrezárt magánhangzónak az ejtési szándék szerinti azonosításában az izolált ejtésű magánhangzókhoz képest. A zöngéindítást a mássalhangzó-környezetnél valamelyest enyhébb, de mégis pozitív hatásnak feltételeztük, mely (az alacsonyabb alapfrekvenciákon) a zöngéindítás nélkül prezentált hangok azonosításához képest valamivel magasabb az ejtési szándék szerinti azonosítási arányokat eredményez.

A vizsgálat eredményei – Smith és Scott (1980) hasonló vizsgálatával ellentétesen – igazolták azt a feltevésünket, mely szerint pusztán a mássalhangzó-környezetben rejlő dinamikus akusztikai kulcsok, azaz a formánsátmenetek nem elegendők ahhoz, hogy a hallgatók az ejtési szándék szerint azonosítsák az alapfrekvencia emelésével egyre változó, és így egyre bizonytalanabb mi-

nőségű magánhangzókat. A zöngéindítás a jelen adatok alapján (melyben csak a magánhangzótér szélső pontjait képező /a: i: u:/ magánhangzókat vizsgáltuk) csak az /u:/ beszédbeli megvalósulásának esetében tűnt az ejtési szándék szerinti azonosítást segítő akusztikai kulcsnak. Ráadásul eredményeink szerint mind a mássalhangzó-környezet, mind pedig a zöngéindítás hatása inkonzisztensnek látszott az izolált ejtésű magánhangzókkal való összevetésében. Az eltérések következtelensége azt jelenti, hogy nemcsak a CVC kontextus, illetve a zöngéindulás pozitív hatását nem sikerült demonstrálnunk, de éppen az elvárásokkal ellenkező példákat is találtunk (a CVC esetében általánosan, azaz minden magánhangzóra a *h* alaphangon, a zöngéindulás esetében pedig a *h* alaphangon az /a:/-ra).

Az eredményeket a vizsgálat kiinduló feltételezésével összhangban a magas alaphangon ejtett (zöngés) hangok spektrális alul-mintavételezettségével és az operaéneklésnek a mássalhangzók képzését érintő sajátosságaival magyarázhatjuk. Mivel az alaphangfrekvencia emelésével e két tényező következtében a magánhangzók akusztikai megvalósulása egyre inkább eltér a hangzók beszédbeli megvalósulásától, így a hangzók minősége egyre bizonytalanabb lesz, tehát nem is lehetséges az ejtési szándék szerinti azonosítás magas alaphangfrekvencián még az egyébként az azonosítást bizonyos más körülmények között segítő akusztikai kulcsok jelenléte esetén sem. Ugyanakkor már önmagában az is kérdés, hogy egyáltalán milyen mértékben segíthetnének a formánsátmenetek akkor, ha a magánhangzók akusztikai szerkezete nem is változna, hiszen a spektrális alul-mintavételezettség nemcsak a vokálisok tiszta fázisát jellemzi, hanem minden olyan beszédhangot és hangrészletet is, melyek akusztikai lenyomatát a hangszalagműködés következtében létrejövő zöngé hozza létre. Ennek megfelelően ugyanis a zöngés mássalhangzók és a bármely két szonoráns között realizálódó formánsátmenetek is alul-mintavételezetté, azaz fonetikailag alulspecifikálttá válnak. A mássalhangzó-realizációk problémájának egy másik oldala az énekelt mássalhangzók artikulációs (és abból következő akusztikai) megvalósítása is. Az operaéneklés (vagy másként a *bel canto* technika) egyik alapvetése ugyanis a magánhangzók időtartamának maximalizálása a mássalhangzók időtartamának rovására. Ennek a mássalhangzó-redukciónak a módját és jellegét azonban, a jelen értekezés szerzőjének ismeretei szerint, kísérletes úton eddig még csak egyetlen alkalommal vizsgálták. Ebben az elemzésben Deme, Grácsi és Jankovics (2013) az énekelt obstruensek zöngéességének alakulására fókuszált. A szerzők egyfelől bemutat-

ták az egyik vizsgált zöngés explozíva (a /b/) rövidülését szókezdő és szózáro helyzetben, másfelől pedig leírták több szókezdő, szó belseji és szózáro explozíva és frikatíva zöngességének megváltozását is. Ezen túlmenően azonban a szerzők arról is beszámoltak, hogy a mássalhangzók címkézésekor gyakori problémát jelentett a vizsgált explozívák felpattanásainak elmaradása (vagy legalábbis detektálhatatlansága a spektrogramon és az oszcillogramon), illetve a réshangok turbulens zöreijösszetevőinek alacsony intenzitása. Az obstruensek lenizációja, illetve ennek megfelelően az egyes akusztikai összetevőinek csekély intenzitása vagy hiánya azonban nemcsak magukra a mássalhangzókra, illetve azok azonosíthatóságára van hatással, hanem a magánhangzók ejtésére, illetve a koartikulációra, és annak akusztikai lenyomatára, a koartikulációs formánsátmenetekre is. Ha ugyanis a magánhangzót megelőző explozíva felpattanása/frikatíva zörejes eleme elmarad, a formánsátmenetek sem olyan módon jelennek meg (vagy nem is jelennek meg egyáltalán) a magánhangzó akusztikumában (és az azt vizualizáló spektrogramban), mint egy gondosan artikulált explozíva/frikatíva esetén, hiszen a nyelv feltehetőleg csak részben teszi meg (vagy esetleg egyáltalán nem) a felpattanás/réses zörej létrehozásához egyébként bejárt útvonalat a szájüregben. Mindez pedig egyúttal azt is jelenti, hogy az éneklésben a nem gondosan artikulált obstruensek környezetében elhelyezkedő magánhangzókban és a mássalhangzó-magánhangzó átmenetekben a beszédben ejtett beszédhangoknál kevesebb dinamikus információ rejlik, tehát kevesebb a percepciót elméletben támogatni képes „akusztikus mankó” is. Okfejtésünkéből egyértelműen következik, hogy a koartikulációs formánsátmenetek hatásának vizsgálatához az énekelt mássalhangzók realizációjának és esetleges redukciójának kutatása is szükséges. Addig ugyanis nincs okunk feltételezni, hogy a mássalhangzó, illetve a koartikulációs formánsátmenetek bármilyen befolyással bírhatnak a percepció működéseire az énekelt magánhangzók esetében, míg az sem tisztázott, hogy egyáltalán maga a konszonáns és a koartikuláció, valamint a koartikulációs formánsátmenetek megjelennek-e, és így elérhetőek lesznek-e az akusztikumban, illetve a percepció számára. Az éneklésben realizálódó mássalhangzók akusztikai és lehetőség szerinti artikulációs vizsgálata a jelen értekezés szerzőjének jövőbeni terveihez tartozik.

Mindemellett fontos kiemelnünk azt is, hogy adataink kimerítő értelmezéséhez szükséges lesz a különböző képzéshelyű mássalhangzók (tehát nem csak a bilabiális, de az alveoláris és veláris felpattanók) vizsgálata is, melyek képzéshelyük (illetve eltérő frekvenciaszerkezetük) követke-

tében eltérő formánsátmenetekkel realizálódnak a CVC szekvenciákban. Ezt, tehát a képzéshely okozta különbségeket szintén következő kutatásainkban tervezzük vizsgálni.

A magánhangzó-azonosítási arányok kondíciók szerinti összehasonlításakor találtuk meg a mássalhangzó-környezettel összefüggésbe hozható egyetlen percepciós tendenciát: a magánhangzók ugyanis a CVC kondícióban mutatták a legszéttartóbb azonosítási tendenciákat. Az /a:/ esetében a CVC kondícióban tapasztaltuk a legmagasabb, míg az /i:/ esetében a legalacsonyabb arányokat; a két, a nyíltsági fok mentén eltérő hangzó észlelési tendenciái ebben a kondícióban tértek el a legnagyobb mértékben. Bár e magánhangzók frekvenciaszerkezetbeli eltérései miatt a percepciós tendenciákban különbségeket várunk, mégis a fent említett eredmény jelentőségére pusztán a jelen vizsálat adatai alapján nem következtethetünk.

Kiemelendő, hogy a produkció módja (az ének mód, illetve a beszéd mód váltása) két okból is az észlelést befolyásoló tényezőnek tűnik. Egyfelől az énekelt mintákkal szemben a beszédben ejtett hangzók több esetben is (az összesített eredményekben, illetve az /u:/ esetében is) mutatták a feltételezett, de az éneklésre nem igazolódott tendenciát, azaz a mássalhangzó-környezet és a zöngéindulás pozitív hatását. Másfelől pedig az /a:/ és /i:/ esetében a beszélve ejtett realizációk mutatták a legmagasabb azonosítási arányokat (alaphangmagasságtól függetlenül) még azokhoz az énekelt alaphangmagasságokhoz képest is, melyek a beszéd átlagos f<sub>0</sub>-értékéhez közeliak voltak.

Amint azt összegző alfejezetünk elején már megemlítettük, a jelen kutatás eredményei ellentmondanak a mássalhangzó-környezetet vizsgáló egyetlen korábbi tanulmányban (Smith–Scott 1980) olvashatóknak. Megválaszolandó kérdés, hogy mivel magyarázható az ellentmondás. Amint arra már több ponton is utaltunk értekezésünkben, Smith és Scott (1980) eredményei az adatokat befolyásoló változók nem szigorú kontrollálása következtében nehezen (vagy talán egyáltalán nem) interpretálhatók. A szerzők ugyanis az általuk vizsgált két kondícióval két jelentősen eltérő percepciós működést vetettek össze, melyek ráadásul jelentősen eltérő számú válaszlehetőséget engednek meg az azonosítási feladat során. Míg a mássalhangzó-környezetben a kísérleti személyek értelmes szavakat kellett, hogy azonosítsanak, ami gyakorlatilag négy lehetőség közti választást jelentett, addig az izolált ejtésben magánhangzókat kellett felismerniük, tehát utóbbi esetben mindösszesen az elemi hallási észlelés és az alsóbb szintű (akusztikai elemzésre

hagyatkozó) beszédészlelési folyamatokra támaszkodhattak a válaszadáskor, mely során így az angol nyelv összes szóbajöhető magánhangzója (azaz legalább 10 lehetőség) közül kellett választaniuk. Smith és Scott (1980) vizsgálatában ily módon vegyült a mássalhangzó-környezet és a magánhangzó-azonosítás problémájának két, egymástól elkülönítendő, hiszen alapvetően eltérő aspektusa, ezért a vizsgálat eredményei valójában sem az egyik – a formánsátmenetek mint akusztikus kulcsok hatásával kapcsolatos –, sem pedig a másik – a magasabb szintű perцепciós működések aktivációjával kapcsolatos – problémakörre nem engednek következtetéseket levonni. A jelen alfejezetben bemutatott vizsgálat ezzel szemben az itt említettek közül kizárólag az első kérdésre fókuszált, és a jól kontrollált feltételek mellett kapott eredményeinkből azt az ezen vizsgálatot megelőző perцепciós tesztjeinkkel is összeférhető következtetést vonhattuk le, hogy a mássalhangzó-környezet nem segíti a magas alaphangfrekvencián énekelt magánhangzóknak az ejtési szándék szerinti azonosítását a koartikulációs formánsátmenetek révén.

A következő alfejezetben bemutatásra kerülő kísérletünkben – bizonyos értelemben tehát a fent ismertetett kísérlet folytatásaképp – a mássalhangzó-környezet hatásának másik aspektusát, a magánhangzók értelmes szavakba ágyazottságának hatását teszteljük és vetjük össze a mássalhangzó-környezet mint dinamikus akusztikai kulcsok tárának hatásával az énekelt magánhangzók észlelésében. Reményeink szerint e két kérdés szétválasztása és összehasonlítása megerősíti azt a javaslatunkat, mely szerint a kontextus hatásának kérdése az észlelésben valójában két, egymástól élesen elkülönítendő fogalmat jelöl, és magyarázattal szolgáltató arra is, miért kaptunk a jelen kísérletben Smith és Scott (1980) eredményeitől jelentősen eltérő eredményeket.

#### **2.4. A mássalhangzó-környezet hatása az énekelt magánhangzók észlelésére: a magánhangzók azonosítása értelmes és értelmetlen szavakban**

A magas alaphangfrekvenciákon énekelt magánhangzókat sajátos artikulációs és akusztikai jellemzőik miatt bizonyos értelemben fonetikailag alulspecifikáltnak tekinthetjük. Az alulspecifikáltság két aspektusa a) az artikulációs és akusztikai célkonfigurációk megváltozása, az  $F_1 : f_0$  formáns-hangolás megjelenése a magánhangzóejtésben magas alaphangfrekvenciákon (azaz az  $F_1$  formáns az

énekhang alaphfrekvenciájára hangolása akkor, ha a megemelt  $f_0$  egyébként magasabb lenne, mint a magánhangzónak a beszédben jellemző  $F_1$  formánsa), valamint b) a magánhangzó-realizációk spektrumának alul-mintavételezettsége (azaz a spektrális tér kitöltöttségének csökkenése). Mindkettő egyre fokozódik az alaphfrekvencia növelésével. A fonetikai alulspecifikáltság miatt kérdés, hogy miként és milyen alaphfrekvenciáig ismerhetők fel az énekelt magánhangzók az ejtési szándék szerint – valamint az is, hogy létezhetnek-e olyan „percepciós mankók”, melyek az ejtési szándék szerinti azonosítást támogatják. Amint azt már az értekezés korábbi pontjain bemutattuk, a beszédpercepció szakirodalmára támaszkodva az énekelt magánhangzók vizsgálatában sem ritka, hogy egyes kutatók a magánhangzók azonosítását segítő körülményként a mássalhangzó-környezet hatását feltételezik. Ugyanakkor azt is láttuk, hogy e feltételezés az éneklésre vonatkoztatása valójában nélkülözi az empirikus megalapozottságot, hiszen az egyetlen, a kérdést vizsgáló kísérlet eredményei a kísérlet módszertani megoldásai (és a confound hatás) miatt csak korlátozottan értelmezhetők. A mássalhangzó-környezet hatásának kérdése az énekelt magánhangzók azonosításában tehát nem tekinthető egyértelműen tisztázottnak.

Az előzőekben már ismertettük azt a feltételezésünket, mely szerint a mássalhangzó-környezet hatását vizsgáló vagy feltételező szakirodalom problémájának egyik fő forrása a mássalhangzó-környezet, illetve a kontextus fogalmának pontatlan meghatározása, használata. Azt tapasztalhatjuk ugyanis, hogy a feltételezések alapvetően a mássalhangzók és a kérdéses magánhangzó közti formánsátmenetek mint akusztikus kulcsok szerepére hivatkoznak a magánhangzó-azonosításban, miközben az egyetlen korábbi vizsgálat inkább csak a kontextus mint értelmes szavakba ágyazottság hatásával kapcsolatban enged következtetni. Smith és Scott (1980) kísérlete bár a koartikulációs formánsátmenetek, illetve a dinamikus akusztikai kulcsok vizsgálatát célozza meg, valójában egészen más kérdést vizsgál. Eredményeik ugyanis nem a dinamikus kulcsok, hanem a magánhangzók az értelmes szavakba ágyazottságának pozitív hatását igazolják.

Éppen ezért a jelen értekezésben a két kérdés szétválasztott vizsgálatát, valamint a kétféle hatás és a kétféle percepciós működés összevetését végeztük el: az előző alfejezetben a dinamikus akusztikus kulcsok, a jelen alfejezetben pedig az értelmes szavakba ágyazottság hatását teszteltük Smith és Scott (1980) eredményeiből kiindulva. Smith és Scott (1980) vizsgálata bizonyította, hogy a magánhangzók értelmes szavakba ágyazottan magasabb arányban ismerhetők fel az ejtési

szándék szerint akkor, ha a szóba jöhető értelmes szavak száma alacsonyabb, mint amit az értelmetlen szavak és/vagy izolált magánhangzók azonosításának feladata megenged – ez utóbbit ugyanis csak a magánhangzókészlet mérete befolyásolja. Az előző alfejezetben ismertetett kísérletünk szerint ugyanakkor a mássalhangzó-kontextus pusztán a dinamikus akusztikai kulcsok révén nem segíti a magánhangzók az ejtési szándék szerinti azonosítását a magas alaphangfrekvenciájú éneklésben (ahol csak az akusztikus elemzést végző bottom-up folyamatokra támaszkodhat a percepció). A jelen alfejezetben bemutatott vizsgálatban mindezek alapján azt a kérdést járjuk körül, hogy miként hat a magánhangzók felismerésére az, ha – bár értelmes szavakba ágyazottan hallják őket az adatközlők, de olyan módon, hogy – a szóba jöhető értelmes szavak száma nem jelentősen kisebb (közel ugyanolyan nagy számú), mint az értelmetlen szavakban szóba jöhetőké. A vizsgálat tehát egyfelől arra a kérdésre keresi a választ, hogy különbözik-e egymástól a mássalhangzó-környezetnek mint a percepciót segítő akusztikai kulcsok tárának és a mássalhangzó-környezetnek mint a magasabb észlelési folyamatokat és a nyelvi tudást is aktiváló körülménynek a hatása.

A jelen vizsgálat ugyanakkor alkalmat szolgáltat arra is, hogy teszteljünk egy másik, az énekelt magánhangzók azonosításával kapcsolatban gyakran felmerülő kérdést is. Percepciós tesztjeink során gyakran találkozunk énekelni nem tudó adatközlőkkel, akikkel a kutatás kapcsán az éneklésről és a beszédképzés folyamatairól is beszélgetünk. Ilyen alkalmak során számos a beszéd- és énekhangképzést érintő naiv elképzelést ismerünk meg. Ezek közül egy (viszonylag gyakran előkerülő) intuitív feltételezés szerint az énekesek a nem énekeseknél a percepció tekintetében is jobban teljesítenek, azaz nem csak, hogy képesek beszédet képezni, beszédhangokat ejteni magas alaphangfrekvencián, de ugyanezen beszédhangoknak az ejtési szándéknak megfelelő azonosításában is jobban teljesítenek. Az adatközlők általában arra alapozzák ezt a feltevésüket, hogy úgy vélik, a nem énekesekkel ellentétben az énekesek pontosan tudják, hogy milyen artikulációs-akusztikai módosításokat hajtanak végre egy-egy magánhangzó ejtésekor a magas és szép (azaz homogén hangszínezetű) énekhang létrehozásának érdekében, éppen ezért arra is képesek, hogy a módosításokról szóló tudásuk alapján visszakövetkeztessenek az eredetileg ejteni szándékolt hangra. Érdekes módon ez a talán naivnak ható elképzelés jól egybeesik a beszédpercepció szakirodalmából ismeretes motoros elmélettel (vö. a jelen értekezés *1.2.2.1. fejezetét*). Emlékeztetőül, a



Liberman és Mattingly nevéhez fűződő elmélet központi kérdése, hogy hogyan birkózik meg az emberi beszédészlelés a beszéd szegmenseinek variabilitásával, mely a beszélők között és beszélőn belül sem állandó megvalósításokból fakad. Az elmélet szerint az észlelés alapvetően nem az akusztikai mintázatok felismerésén, hanem (az akusztikai jelekből kikövetkeztethető) artikulációs gesztusok azonosításán múlik. Az elmélet követői úgy vélik ugyanis, hogy az agy nem a hangok akusztikai lenyomatát, hanem a beszédhangok ejtéséhez szükséges motoros parancsokat (artikulációs gesztusokat) reprezentálja (mégpedig invariáns módon), és az észlelt akusztikai jel is, gyakorlatilag automatikusan, ilyen információegységekké transzformálódik az észlelési folyamatok során (Liberman–Mattingly 1985). A motoros elméletből következően a megvalósítás stílustól, beszédhelyzettől függően változatos lehet, de mégis minden megvalósításban benne rejlenek a változatlan alapegységek, mégpedig a produkcióhoz szükséges artikulációs gesztusok, amelyek az észlelést is lehetővé teszik. Ez az elmélet tehát bizonyos tekintetben megerősíti az imént ismertetett intuitív elképzelést, mely szerint az énekesek saját artikulációs gyakorlatukból kifolyólag a nem énekes hallgatónál sokkal inkább képesek a jelentős akusztikai alulspecifikáltsággal megjelenő hangzók azonosítására. A motoros elméletet sokan sok okból vitatják<sup>19</sup>, és a jelen értekezésben nem áll szándékunkban állást foglalni e vitában. Mégis, ha azt a feltételezést nem is fogadjuk el, hogy az artikulációs gyakorlat közvetlenül összeköttetésben volna a percepció mechanizmusával, az továbbra is kérdés marad, hogy az énekesek artikulációs tudatossága működésbe léphet-e (és ha igen, pozitívan járul-e hozzá) az énekeseknek a magas alaphékvencian énekelt magánhangzók azonosításában nyújtott teljesítményéhez. Kérdés tehát, hogy jobban teljesítenek-e az énekesek az énekelt magánhangzók az ejtési szándék szerinti azonosításában, mint az éneklésben nem gyakorlott észlelők. Ismereteink szerint ezt a felvetést korábban még nem vizsgálta tudományos kísérlet.

Az alfejezetben ismertetett kísérletek célja tehát kettős. Az első cél az eltérő percepció működéseket feltételező feladattípusokban (azaz az értelmes és értelmetlen szavakban) tapasztalható

---

<sup>19</sup> Egyrészt a motoros elmélettel szemben állnak az akusztikum elsődlegességét hangsúlyozó, illetve az agyi reprezentációk variabilitását feltételező fonetikai teóriák (utóbbira híres példa Lindblom 1990-ben közölt H&H elmélete, vö. még a jelen dolgozat 1.2.2.2. fejezete). Másrészt pedig idegtudományi vizsgálatok is cáfolják a motoros teóriát: Hickok (2010) például, aki az agyi aktivitás és a tükroneuronok vizsgálata alapján azt feltételezi, hogy az észlelés közben tapasztalható motoros agyi aktivitás valójában a szenzoros területekről terjedő aktiváció eredménye, aminek célja nem az észlelési folyamatok segítése, hanem a megszólalásra, a beszédprodukcióra való felkészülés.

magánhangzó-azonosítási tendenciák összehasonlítása, ezzel is tovább árnyalva a mássalhangzó-kontextus hatásának kérdését a magas alaphfrekvencián énekelt magánhangzók észlelésében. Ed-digi akusztikus és észlelési vizsgálataink alapján azt feltételezzük, hogy nem hatékonyabb az ej-tési szándéknak megfelelő magánhangzó-azonosítás az értelmes szavakban akkor, ha az értelmes és értelmetlen szavakban ejtett magánhangzók azonosítását nem jelentősen eltérő (de akár nem is szükségyszerűen ugyanannyi) válaszlehetőséget megengedő feladatokban vetjük össze. Feltevé-sünk szerint ugyanis a kontextus pozitív hatása nem a koartikulációs formánsátmenetekben, ha-nem az értelmes szavak által korlátozott válaszlehetőségekben rejlik, mely hatás közel kiegyenlí-tett számú válaszlehetőségek esetén erősen lecsökken (vagy teljesen meg is szűnik). A kísérletünk másik célja az énekes és nem énekes hallgatók teljesítményének összehasonlítása a magánhang-zó-észlelés tekintetében. Elfogadva a motoros elmélet egy lehetséges értelmezéséből kiinduló fel-tételezést, mely szerint az énekeseknek a saját artikulációs működéseikkel kapcsolatos tapasztala-ta vagy tudása a beszédpercepció valamely szintjén képes beépülni az észlelési folyamatokba, és így percepciós előnyt jelent, azt feltételezzük, hogy az énekesek nagyobb arányban képesek az ejtési szándéknak megfelelő hangzóazonosításra, mint a nem énekes hallgatók.

#### **2.4.1. Anyag, módszer, kísérleti személyek**

A hipotézisek vizsgálatára egy két ülésből álló percepciós tesztet terveztünk, melyet 20 „é-ne-kes” (26 és 45 év közöttiek, átlagos életkoruk 31 év) és 20 „nem énekes” (26 és 52 év közöttiek, átlagos életkoruk 29 év) adatközlővel vettünk fel. Mindannyian egészséges, ép hallású, magyar anyanyelvű beszélők voltak. Az énekesek csoportjában 15 profi szoprán (nő), 2 profi tenor és 3 profi bariton (férfi) képviselte a magas alaphfrekvenciákhoz szokott, klasszikus nyugati típusú hangképzést tanult operaénekeseket. A „nem énekes” csoportban fonetikailag és énektechnikailag képzetlen, „naiv” adatközlők szerepeltek.

A percepciós kísérlethez két stimuluskészletet állítottunk elő, majd a stimulusokat egy 29 éves profi szoprán énekesnő ejtésében rögzítettük csendesített szobában, gömbkarakterisztikájú mikro-fon segítségével. Az egyik készlet az értelmes szavakat tartalmazta („Szavak” kondíció), míg a másik készlet az értelmetlen szavakat („Álszók” kondíció). Mindkét készlet CVC szerkezetű szekvenciákból állt, melyekben a magánhangzókészlet maximális kihasználása, valamint a

koartikulációs hangátmenetek egységesítése miatt az /ɒ a: ɛ i: ø: y:/ magánhangzók szerepeltek alveoláris képzéshelyű mássalhangzók között. A „Szavak” kondícióban a vizsgált hangokat az *szVr* mássalhangzóvázba ágyaztuk (pl. *szár* /sa:r/), míg az „Álszók” kondícióban *dVr* kontextust használtunk (pl. *dár* /da:r/). A stimulusokat az énekes hat, egymástól (a temperált hangolás szerint) egyenlő távolságra lévő alapfrekvencián énekelte fel állandó alapfrekvencián, a számára kényelmes hangerősséggel (és vibratóval), 175 Hz és 988 Hz között (az *f, h, f', h', f'', h''* zenei hangokon, l. 6. táblázat), ezzel gyakorlatilag teljesen lefedve a szoprán hangosztály alaphangtartományát. Az énekelt stimulusok mellett a hangsorokat az énekes beszélt ejtésében is rögzítettük. Így összesen ( $7 \text{ f0} \times 2 \text{ kondíció} \times 6 \text{ magánhangzó}$ ) 84 teszthangsort vizsgáltunk.

A percepció tesztet egy, a Praat programban (Boersma–Weenink 2009) írt és futtatott szkripttel vezéreltük. A teszthangsorokat kondíciók szerinti bontásban összesen 42 disztraktor szóval és álszóval együtt két részben mutattuk be a résztvevőknek. Minden résztvevő minden hangsort kétszer hallott, de minden résztvevő más-más (randomizált) sorrendben hallotta a teszthangsorokat, így az esetleges fáradás következtében megjelenő hatások eloszlottak a stimulusok között. Az első részben minden esetben a „Szavak” kondíció stimulusait tartalmazó tesztet végeztük el, majd pár perc szünet beiktatása után ezt az „Álszók” kondíció elemeit tartalmazó rész követte. A résztvevők feladata mindkét tesztfázisban ugyanaz volt: ki kellett választaniuk a képernyőn ortografikus alakban megjelenő kilenc magyar magánhangzó közül azt, amelyiket hallani vélték. A válaszlehetőségek a hosszú (vagy fonémakategória-váltás nélkül hosszan ejthető) magyar magánhangzók voltak: /ɒ a: ɛ e: i: o: ø: u: y:/. Az adatközlőknek szóló instrukciók azonban némileg eltértek a két tesztben. Míg az első részben felhívtuk az adatközlők figyelmét, hogy értelmes magyar szavakat fognak hallani, addig a másodikban nyomatékosan megkértük őket, hogy kifejezetten az elhangzó magánhangzó minőségére figyeljenek, ugyanis értelmes és értelmetlen hangsorok egyaránt előfordulhatnak a bemutatott stimulusok között. Mivel a képernyőn bemutatott választható magánhangzók mindkét tesztben ugyanabból a kilenc magánhangzóból álltak, így az első fázisban olyan lehetőségeket is megjelenítettünk, melyek ott nem tartoztak a legitim megoldások közé, mégpedig az /e: o: u:/ hangokat. (Tekintve, hogy az /o: u:/ magánhangzók az „Álszók” *dVr* mássalhangzóvázában is értelmes szót alkotnának, ezért ezek vizsgálatát kizártuk a jelen kísérletből, az /e:/ hang pedig a „Szavak” *dVr* mássalhangzóvázában is értelmetlen szót al-

kotott volna, így ebben a kísérletben ezért nem vizsgáltuk). Ez a megoldás tudatos döntés eredménye volt, mert fontosnak tartottuk fenntartani annak a lehetőségét, hogy a résztvevők szabadon „hibázhassanak”. Más vizsgálatokban kapott eredményeink alapján ugyanis feltételezhető például, hogy az /i:/ már igen alacsony alapfrekvencián is elveszíti eredeti perceptuális minőségét, és a hallgatók /e:/ fonémaként kategorizálják (l. 2.1. és 2.5. *alfejezet*). Érdekesnek tartottuk tehát megfigyelni, hogy mi történik, ha a feladatutasítás (mely szerint az /e:/ nem legitim válasz) és a hallott magánhangzó perceptuális fonémakategóriája (mely feltételezhetően már alacsony alapfrekvencián is az /e:/ hang megjelölésére sarkallná a résztvevőket) konfliktusba kerül. A kísérlet e sajátosságának következtében azonban az eredmények taglalásakor mindig szem előtt kell tartanunk két fontos szempontot. Az egyik szempont az, hogy az egyes magánhangzókra kapott azonosítási arányokat ne abszolút módon értelmezzük, hanem annak is a szem előtt tartásával, hogy a „Szavak” kondícióban egy-egy magánhangzóválasz csekély mértékben, de nagyobb valószínűséggel **jelenthetett** meg. Ha a résztvevő a fent említett három magánhangzót kizárta a válaszlehetőségek közül, akkor ez a valószínűség a „Szavak” kondícióban 17%, míg az „Álszók” kondícióban 11% volt. Azonban ha az adatközlő a szólista figyelmen kívül hagyásával döntött (tehát az /o:/ u:/ magánhangzókat is legitim válasznak tekintette), a valószínűségek rendre 13% és 11% voltak. Ha pedig az adatközlő a „szabályok betartásával” végezte el a tesztet, a valószínűségek egyenlők (11-11%) voltak. A másik szempont mindennek megfelelően pedig az, hogy nem csak az ejtési szándéknak megfelelő válaszokra és a „helyes tippelésre”, de az ejtési szándéktól eltérő azonosítások bizonyos tendenciáinak megjelenésére is nagyobb esély lehetett a „Szavak” kondícióban. Ezáltal a kísérletben valójában tovább megyünk annál, mintsem hogy csak az értelmes és értelmetlen szavakban megmutatkozó magánhangzó-azonosítás jellemzőit vetnénk össze (erre elegendő lenne, ha 6-6 válaszlehetőséget hagynánk a résztvevőknek). Demonstrálni kívánjuk ugyanis azt is, hogy a két vizsgált feladat közti alapvető különbség a válaszlehetőségek számában rejlik. Eszerint (a valószínűségek természetéből fakadóan) nagy különbségeket okoz, ha kisszámú válaszlehetőség áll szemben nagyszámú válaszlehetőséggel (ti. a magánhangzórendszer összes magánhangzójával), ahogy Smith és Scott (1980) kutatásában, de jóval kisebb különbséget, illetve a különbségek tökéletes eltűnését eredményezi már az is, ha az összevetett csoportokban azonos vagy hasonlóan nagyszámú válaszlehetőség van.

A bemutatások során a mássalhangzó-kontextust vizuálisan is megjelenítettük a képernyő felső részén, így maximalizálva a koartikulációs hangátmenetekből eredhető pozitív hatást. A „Szavak” kondíció tesztfázisában emellett az összes lehetséges értelmes magyar szót is felsoroltuk ugyanitt, tehát az adatközlőket minden bemutatás alkalmával emlékeztettük arra, melyek a legitim választások. A teszteket csendes körülmények között vettük fel, a hanganyagokat zárt fejhallgatón keresztül hallották a résztvevők.

Az eredményeket alaphékvenciánként az összes magánhangzóra összesítve és az egyes magánhangzókra külön-külön is elemeztük a résztvevők két csoportjában, valamint megvizsgáltuk a nem az ejtési szándék szerint azonosított hangzók esetében fellelhető tendenciák típusait is a két kondícióban. Adatainkon az R programmal (R Core Team 2013) végeztünk statisztikai elemzéseket: előbb Shapiro–Wilk-próbával ellenőriztük a normális eloszlás kritériumát, majd normális eloszlás esetén parametrikus teszteket ( $t$ -próba, párosított  $t$ -próba, ANOVA, ismételt mérés ANOVA), nem normális eloszlás esetén pedig nem parametrikus teszteket (Mann–Whitney U-próba) végeztünk.

## **2.4.2. Eredmények**

### **2.4.2.1. Bevezető gondolatok a résztvevők benyomásai alapján**

A teszt felvétele után minden adatközlőt annak összefoglalására kértünk, hogy milyen nehézségűnek találták a két tesztfeladatot, és hogy milyen általános benyomásaik születtek a saját teljesítményükről. A résztvevők mindkét csoportban egyöntetűen állították, hogy a két feladat között nem érzékeltek nehézségi különbséget, valamint hogy a magasabb alaphékvenciákon ejtett hangzók esetében is jól teljesítettek, bár érzékelték, hogy „a magas hangoknál minden olyan á-szerű”. Az énekes csoport tagjai mindehhez még azt is hozzátették, hogy a felvételen szereplő énekes magánhangzóiból kihallani vélték azt a jellegzetes ejtési sajátosságot is, mely tapasztalatuk szerint az elmúlt évtizedekben a magánének-tanításban elfogadott, de mára már túlhaladottnak tekintett megoldásból fakad. Eszerint a tanárok az /i:/ hang helyett /y:/-t, míg az /e:/ helyett /ø:/-t énekeltetnek diákjaikkal, hogy így (tehát az ajakkerekítéses ejtés révén) könnyítsék meg a magas

alaphangok elérését.<sup>20</sup> A válaszok fényében két fontos, a kísérlet eredményeit érintő megfigyelést tehetünk. Egyrészt fontos látnunk, hogy bármilyen artikulációs-akusztikai változáson esnek át a beszédhangok az éneklés során az alapfrekvencia függvényében, a résztvevők minden, a kísérletben elhangzott hangsorban létező magyar magánhangzókat véltek felfedezni, melyek számukra egyértelműen besorolhatónak tűntek egy-egy fonémakategóriába. Másrészt pedig az énekesek kommentárjai arra is felhívják a figyelmet, hogy érdemes lehet megfigyelni, milyen észlelési tendenciákat találunk a „nem énekes” és „énekes” résztvevőcsoportokban (és egymással összevetve) az /i: y: e: ø:/ hangokra kapott eredmények tekintetében.

#### **2.4.2.2. Az ejtési szándéknak megfelelő magánhangzó-azonosítás az értelmes és az értelmetlen szavakban**

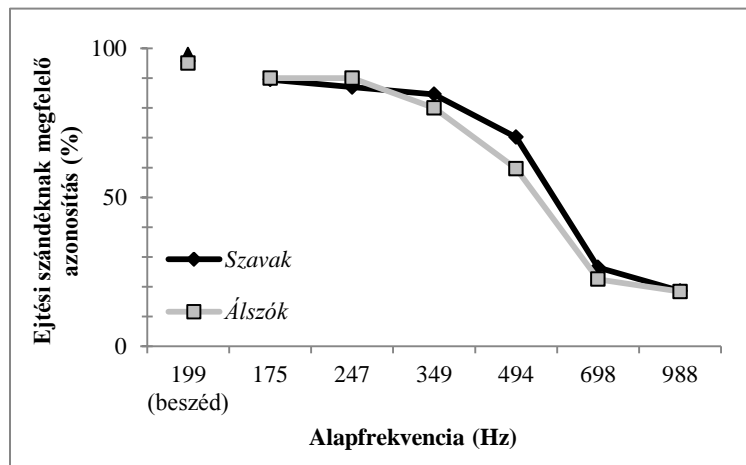
A 22. ábra az ejtési szándéknak megfelelő válaszok az egyes alapfrekvenciáinként jellemző összesített arányát mutatja a „nem énekes” résztvevőknél. Az ábra a következőképpen olvasandó: míg 494 Hz alapfrekvencián a „Szavak” kondícióban a  $(20 \text{ adatközlő} \times 6 \text{ magánhangzó} \times 2 \text{ ismétlés} =) 240$  magánhangzóválasz 70%-a egyezett meg az ejtési szándéknak megfelelő magánhangzóval, addig az „Álszók” kondícióban a 240 magánhangzó-válasznak csak 60%-a volt az ejtési szándéknak megfelelő. Az első, a beszédre kapott észlelési adatokat mutató adatpontot itt is elkülönítve jelenítettük meg, hiszen ez nem csak alapfrekvenciájában, de a produkció módjában is eltért a többi adatunktól.

Az ábrán jól látható, hogy a magánhangzó-azonosítás tendenciái gyakorlatilag megegyeztek a két kondíció esetében. Az apróbb eltérések ellenére a két adatsor közti különbség a páros  $t$ -próba tanúsága szerint statisztikailag nem szignifikáns. A legmagasabb (95% körüli) azonosítási arányokat mindkét kondícióban a beszéd módban látjuk. Ehhez képest az összes, még a beszéd alapfrekvenciájához viszonyítva alacsonyabb alapfrekvenciájú énekelt alapfrekvencia esetében is ala-

---

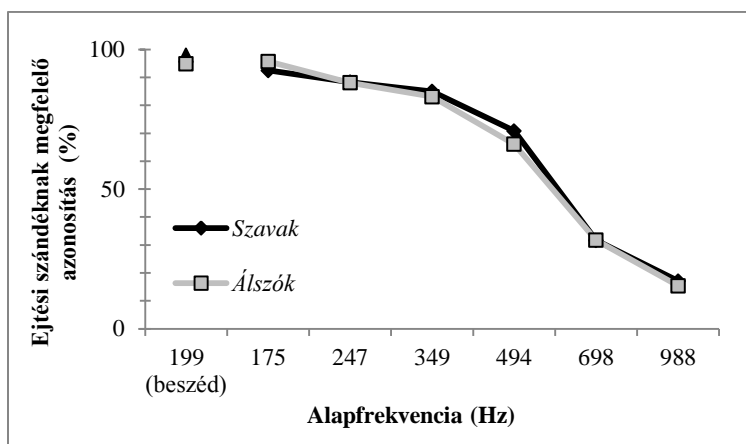
<sup>20</sup> Ennek a stratégiának az oka feltehetőleg abban rejlik, hogy az énekesek igyekeznek kerülni a magasabb gégehelyezettel járó zöngéképzést, mert az könnyen vezethet feszített, préselt fonációhoz (Zenker és Zenker 1960, idézi Pabst és Sundberg 1992), és ezért igyekeznek lesüllyeszteni a gégét. A gége (nyugalmi helyzethez képest) alacsonyabb vertikális pozíciójának eléréséhez pedig Sundberg javaslata szerint hatékony indirekt eszköz lehet a hangzók kerekítéses ejtése (szíves szóbeli közlés). Valóban, az ajakkerekítéses hangzók (pl. az /y:/) esetében a gége jóval (a nyugalmi pozíciónál) lejjebb helyezkedik el, mint az azonos nyelvállásfokú ajakréses párjuk (pl. az /i:/) esetében, ahol a gége a nyugalmi pozíciónál magasabbra emelkedik (l. pl. Hoole–Kroos 1998).

csonyabbak a százalékos arányok. Az általános tendenciák alapján azt mondhatjuk, hogy az ejtési szándék szerinti azonosítás aránya mindkét kondícióban csökkent az alapfrekvencia emelésével, mégpedig fokozatosan, ugyanakkor 500 Hz (azaz a  $h'$  zenei hang) környékén hirtelen esés volt tapasztalható. A két kondíció egyetlen ponton mutat szembetűnőbb különbségeket: 494 Hz alapfrekvencián (a  $h'$  zenei alaphangon) az adatközlők teljesítménye 10%-kal alacsonyabb az álszókban, mint a szavakban.



22. ábra: Az egyes alapfrekvenciákon bemutatott összes stimulus ejtési szándék szerinti azonosítási aránya a két kondícióban a „nem énekes” adatközlőknél

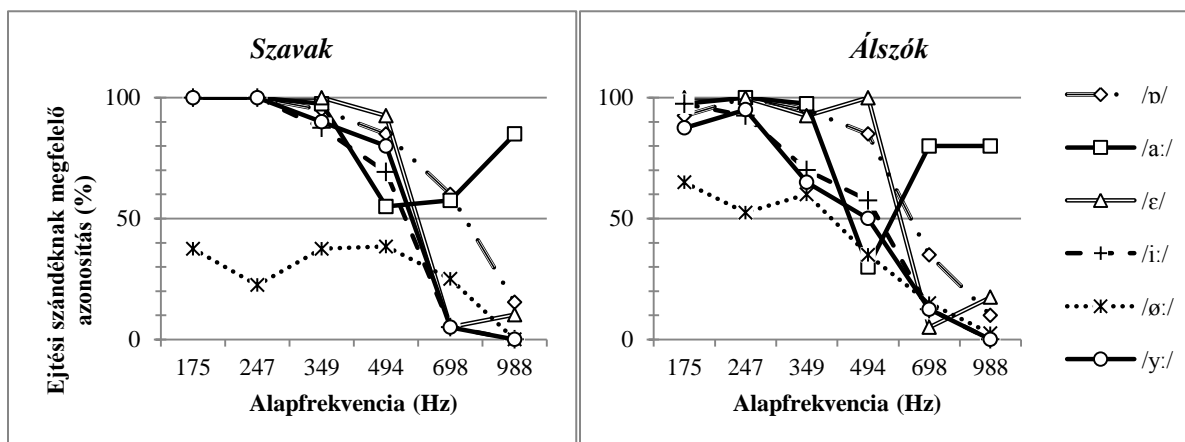
A 23. ábrán az „énekes” résztvevőknél tapasztalt azonosítási arányok láthatók. Az előzőekhez hasonlóan itt is azt találtuk, hogy a két kondícióban kapott válaszok aránya nem tér el egymástól statisztikailag igazolható módon (páros  $t$ -próba). Az azonosításban mutatott magasabb arányokat az adatközlők e csoportjában is a beszéd módban tapasztaltuk. Érdekes módon az „énekesek” esetében nem tapasztaltunk még olyan apróbb különbségeket sem a kondíciók között, mint a „nem énekeseknél”.



23. Az egyes alapfrekvenciákon bemutatott összes stimulus ejtési szándék szerinti azonosítási aránya a két kondícióban az „énekes” adatközlőknél

#### 2.4.2.3. A magánhangzók az ejtési szándéknak megfelelő azonosítása a két kondícióban

A 24. ábrán az ejtési szándéknak megfelelő azonosítási arányokat látjuk magánhangzó-minőségenkénti bontásban, két kondícióban (balra: „Szavak”, jobbra: „Álszók”) a „nem énekes” adatközlőknél. Az eredmények összehasonlítására ismételt méréses vairanciaanalízist alkalmaztunk a *kondíció* faktorról („Szavak”, „Álszók”), valamint a *magánhangzó-minőség* faktorról. A vizsgálat szerint nem mutatható ki a *kondíció* faktor főhatása (within-subject effect alapján) a magánhangzó-minőségek (between-subject effect alapján) figyelembevételével sem.



24. ábra: Az ejtési szándék szerinti magánhangzó-azonosítás tendenciái a „Szavak” (balra) és az „Álszók” (jobbra) kondícióban a „nem énekes” csoportban

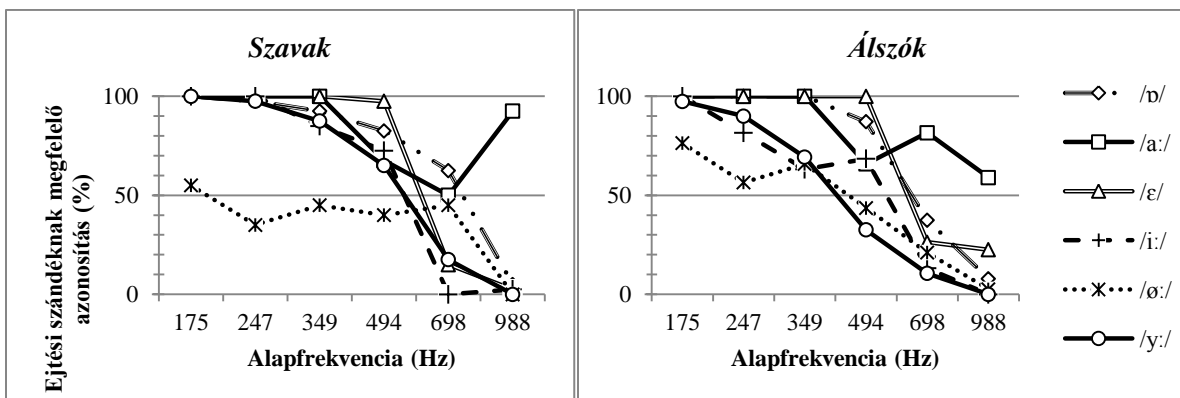


Annak ellenére, hogy a „Szavak” kondícióban az egyes magánhangzók eltalálására akár több esélye is lehetett az adatközlőknek, az /ø:/ esetében általánosan nagyon alacsony, és az „Álszók” kondícióban tapasztalhatóan általánosan alacsonyabb azonosítási arányokat találtunk még a kritikusabbnak tűnő 698 Hz alapfrekvencia ( $f''$ ) alatt is. Az /ø:/ azonosítása a beszédben is elmarad a többi magánhangzótól, ugyanis csak 95%-os, szemben a többi vokálisra jellemző 100%-kal (az ábrán ezek a beszéd módra kapott eredmények nem szerepelnek). A magánhangzónkénti összevetésből az is kiderült, hogy 22. és 23. ábrán látott, a beszédre jellemző alacsonyabb (nem 100%-os) teljesítmény az /ø:/ bizonytalannak tűnő hangszínéből fakad.

A legnagyobb biztonsággal az ejtés szerint azonosított hangzó mindkét kondícióban az /a:/ volt. Az ejtési szándéknak megfelelő azonosításának aránya ugyanakkor egy ponton, 494 Hz alapfrekvencián hirtelen lecsökkent, majd az alaphang további emelésével emelkedett meg ismét. Bár az azonosítás arányaiban talált esés jelentősebbnek tűnik az „Álszók” kondícióban, a *Szavak* kondícióban a százalékok a következő, azaz 698 Hz alapfrekvenciájú alaphangnál alacsonyabbak.

A „Szavak” és „Álszók” tendenciái között talált egy kisebb különbség az elől képzett zárt /i:/ *y:/* eltérő viselkedése a két kondícióban kritikusnak tűnő 698 Hz alatt (ahol az azonosítás általánosságban magasabb arányokat mutat). Az /i:/ és *y:/* ugyanis a „Szavak” kondícióban (tehát a *szír* és *szűr* szavakban) valamivel magasabb arányban értelmeződött az ejtési szándéknak megfelelően – és így hasonlóan a nyíltabb ejtésű /ɒ/-hoz és /ɛ/-hez –, mint az „Álszók” kondícióban (azaz a *dír* és *dűr* szekvenciákban). A két kondíció között ebben a tekintetben a legnagyobb különbséget az /y:/-nél tapasztaltuk, melynek az ejtési szándék szerinti azonosítása 494 Hz alapfrekvencián 30%-kal kisebb volt az „Álszók” esetében.

Az 25. ábra összegzi az énekes adatközlők eredményeit. Ezekre az adatokra (megint csak ismételt méréses varianciaelemzéssel végezve az összehasonlítást) szintén azt találtuk, hogy nem mutatható ki a *kondíció* főhatás a magánhangzó-minőségek figyelembevételével sem. Mind az énekesekre, mind a nem énekes résztvevőkre kapott eredmények azt jelzik, hogy bár az egyes magánhangzókra kapott adatok többé-kevésbé eltérhetnek az egyes alapfrekvenciákon, de a *kondíció* tekintetében nem különülnek el szisztematikusan, ha az összes alapfrekvenciát figyelembe vesszük.

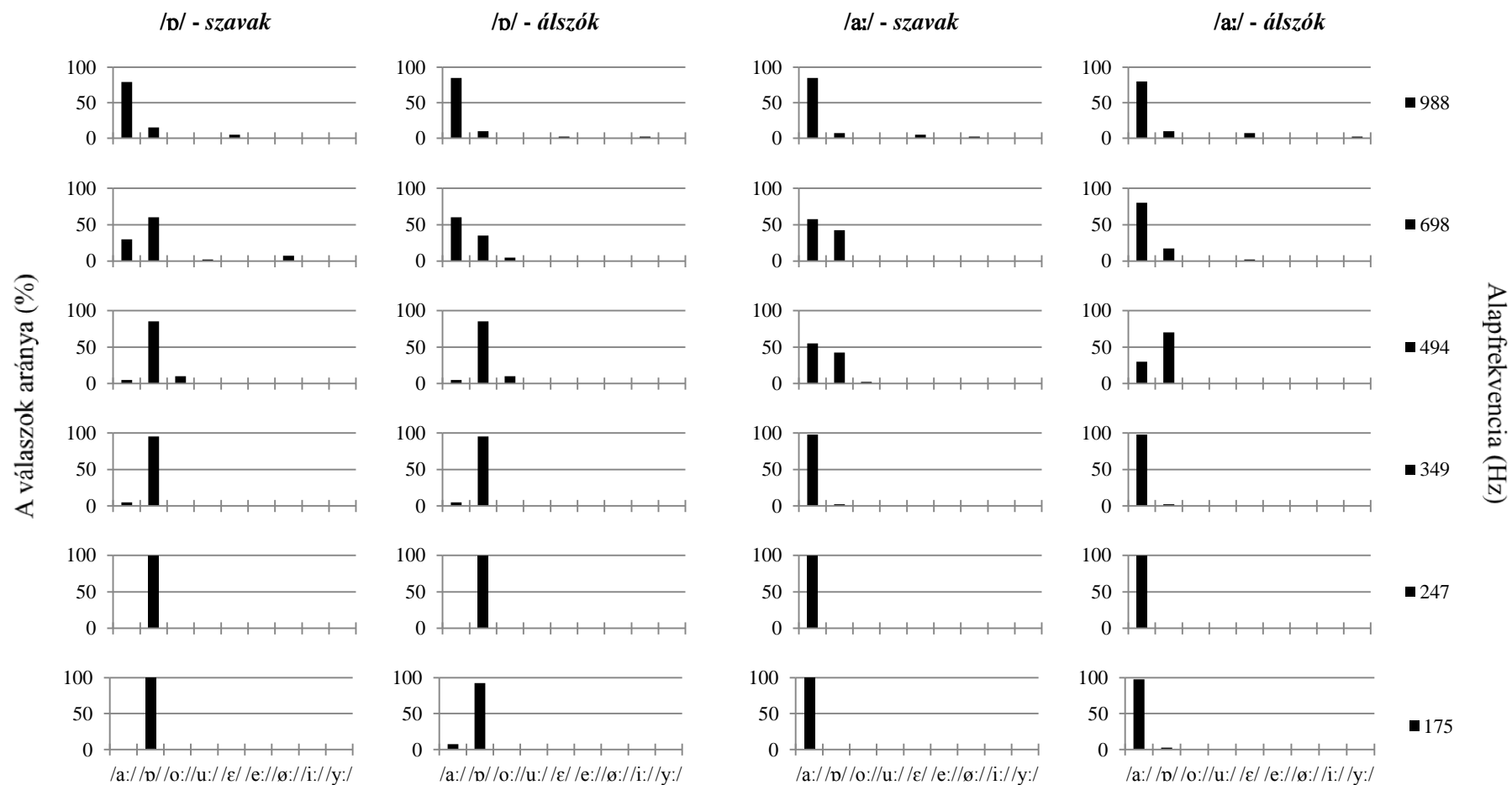


25. ábra: A magánhangzó-azonosítás tendenciái a „Szavak” (balra) és az „Álszók” (jobbra) kondícióban az „énekes” csoportban

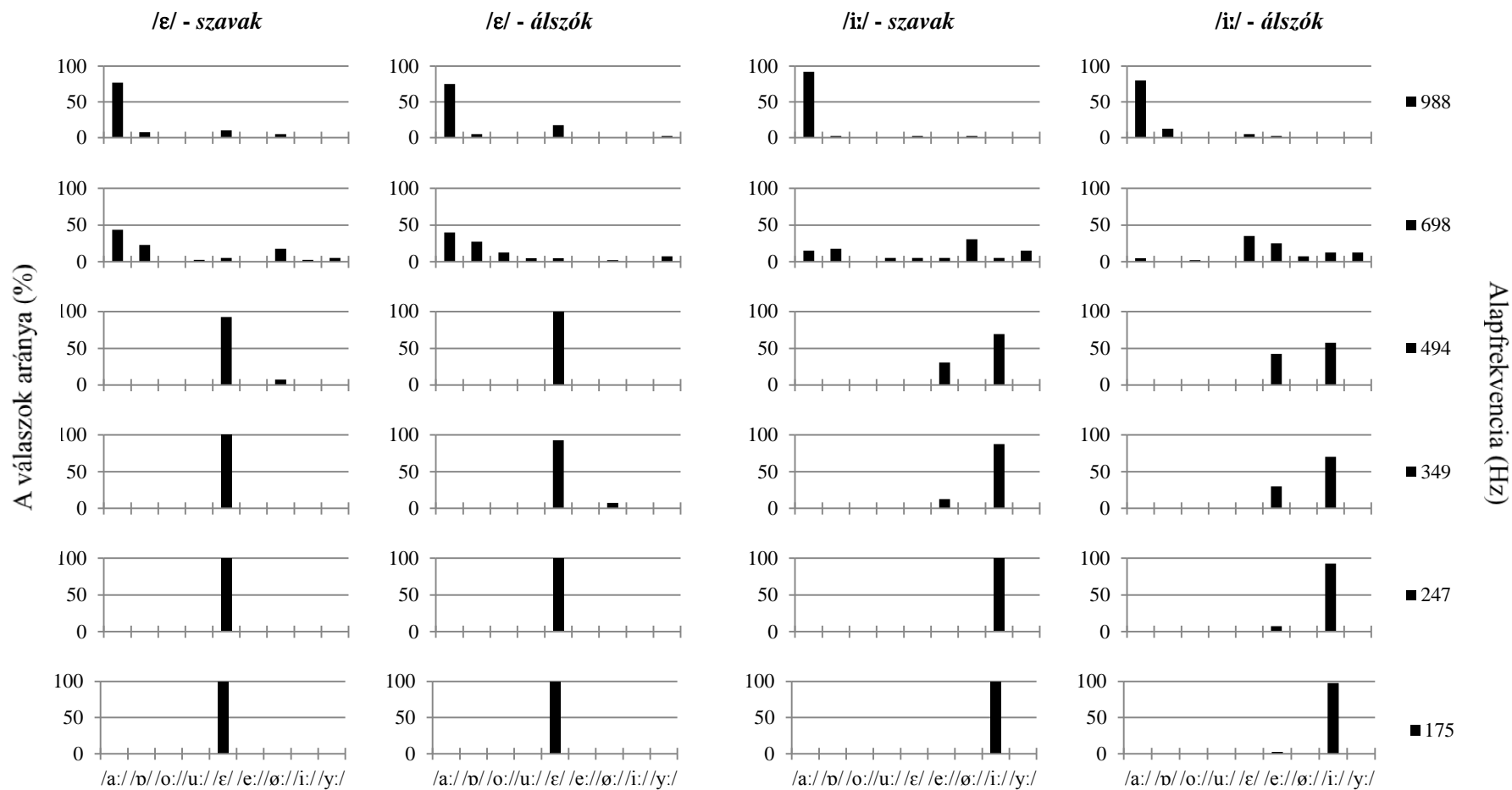
Az /ø:/ sajátos viselkedése, azaz a szavakban és álszókban tapasztalt azonosítási tendenciáinak eltérése az „énekes” csoportban is megfigyelhető volt, ugyanúgy, ahogy az /a:/ ejtési szándék szerinti magas arányú azonosítása. Eltérés azonban, hogy az „énekes” csoportban nem látunk olyan drámai esést az /a:/ esetében az „Álszók” kondícióban 494 és 698 Hz környékén.

#### 2.4.2.4. A tévesztések tendenciái az egyes magánhangzók esetében az alapfrekvencia függvényében

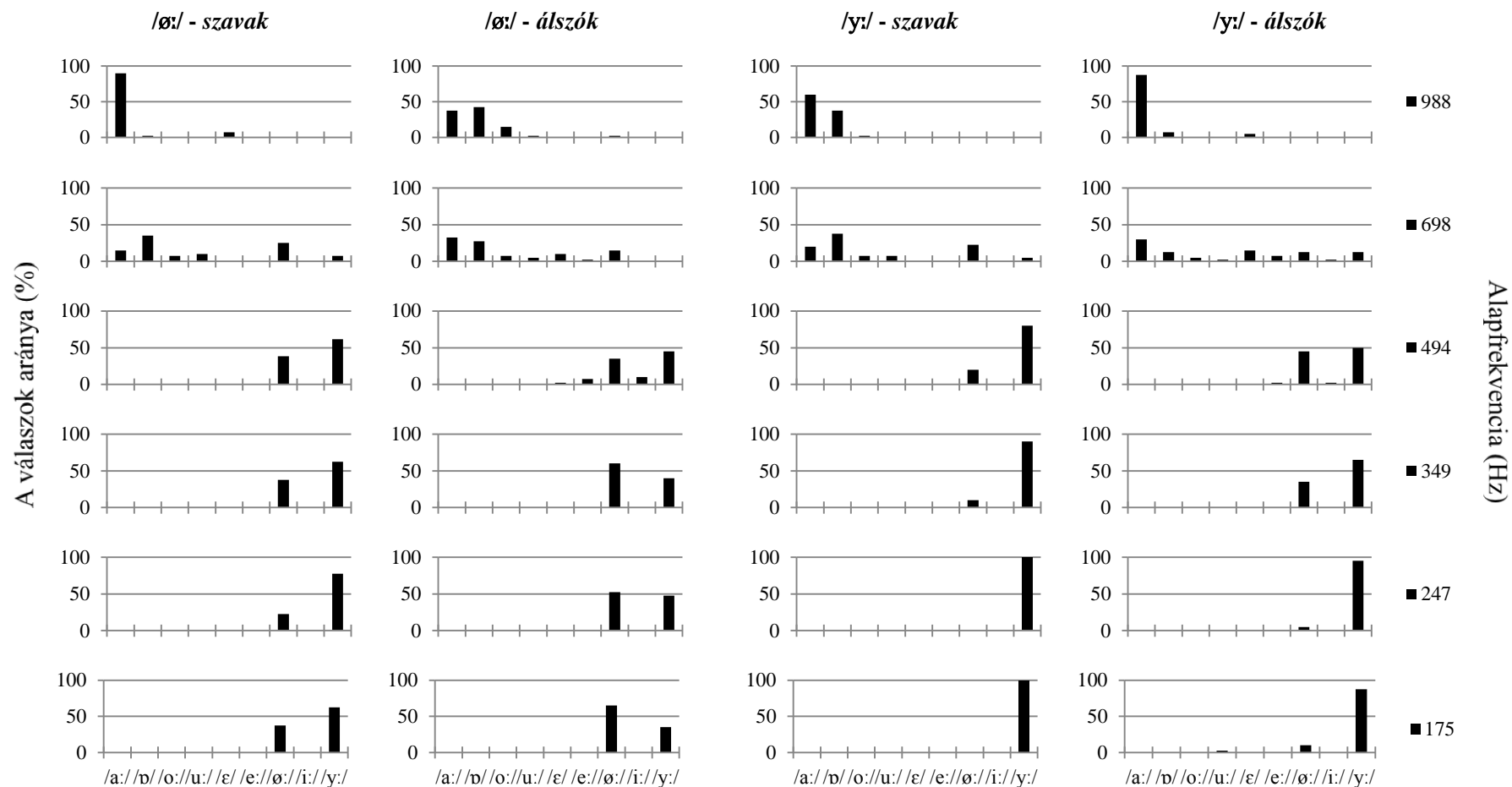
A 26.a, 26.b és 26.c ábrán a „nem énekes” csoportnak az egyes ejteni szándékozott magánhangzókra érkezett válaszait látjuk az alaphang függvényében a két kondícióban. Minden panel az egy-egy alapfrekvencián kapott összes választ (és a kilenc válaszlehetőséget) mutatja (beleértve az ejtési szándéknek megfelelő hangokra kapott válaszokat is) úgy, hogy (a hangmagasságérzetnek megfelelő sorrendben) a legalsó panel a legmélyebb, a legfelső pedig a legmagasabb alapfrekvenciára kapott válaszokat reprezentálja.



26. a. ábra: Az /p a/ magánhangzókra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében a „nem énekes” adatközlőknél.



26. b. ábra: Az /ε i:/ magánhangzókra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében a „nem énekes” adatközlőknél.

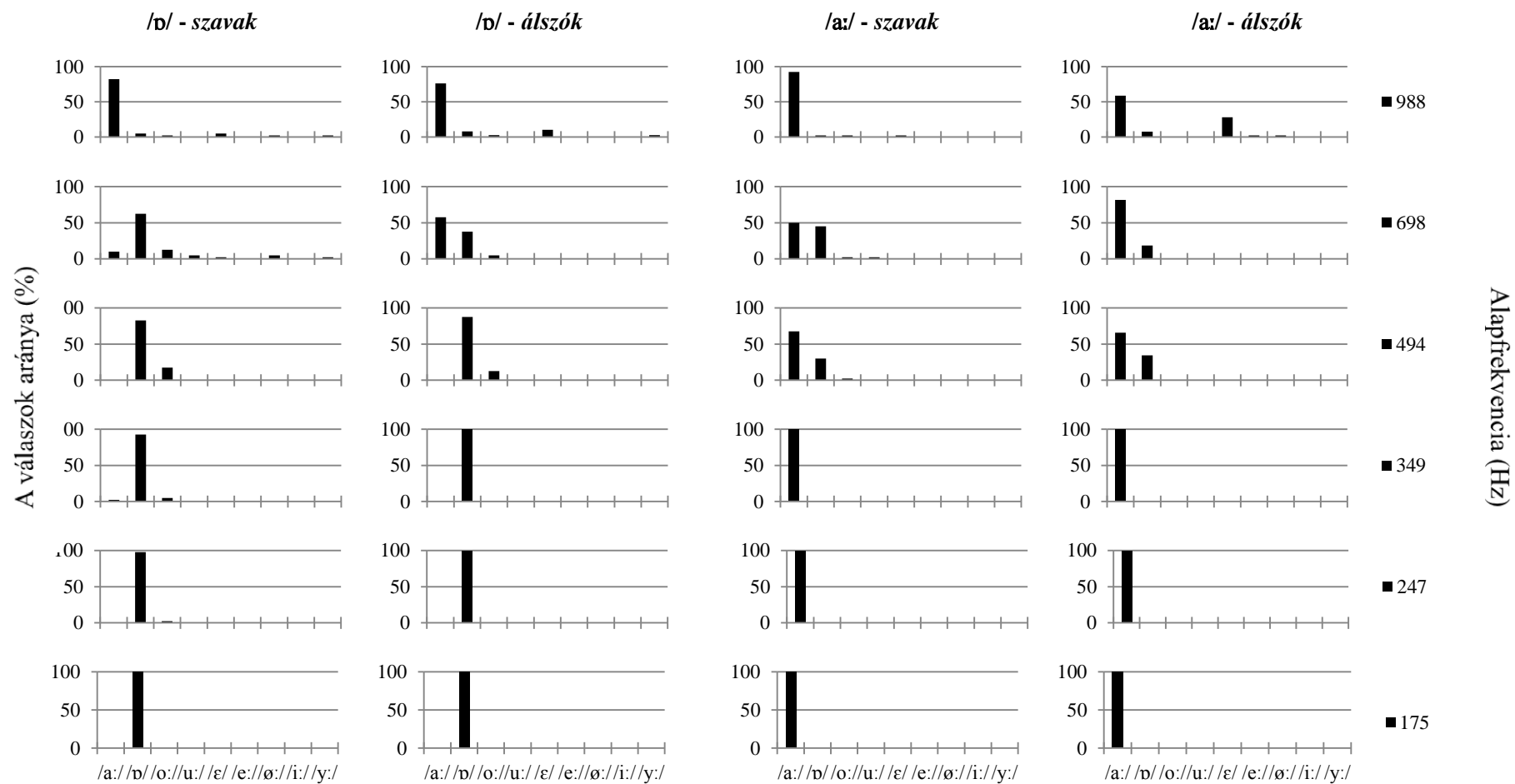


26. c. ábra: Az /ø:/ y:/ magánhangzókra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében a „nem énekes” adatközlőknél.

Az előzőekben azt láttuk, hogy a „nem énekesek” legkisebb arányban mindkét kondícióban az /ø:/ hangot azonosították az ejtési szándék szerint. Az ábra szerint ennek az oka az /ø:/ /y:/-ként történt nagyarányú azonosítása. Az /i:/ és /y:/ esetében magasabb ejtési szándék szerinti azonosítási arányokat találtunk a „Szavak” esetében. A 26 (és különösen a 26.b) ábra tanúsága szerint ez az /i:/ esetében feltehetőleg arra vezethető vissza, hogy míg az „Álszók” esetében 247 és 494 Hz között az /i:/-re egyre nagyobb számban érkeztek /e:/ válaszok, a „Szavak” esetében az /e:/ válaszok megszorodása nem olyan nagy mértékű, hiszen az /e:/ válasz a „Szavak” esetében elvileg illegitim volt. Éppen ezért különösen érdekes megfigyelni, hogy mindellenére mégis milyen nagy arányban jelenik meg az /e:/ válasz a „Szavak” kondícióban is. Az /y:/ esetében a két kondíció közti különbséget 698 Hz alatt egyértelműen az /ø:/ válaszok arányának eltérése okozza (ez mindkét kondícióban legitim válasznak minősült). Az adatokból az is kiderül, hogy az egyébként általánosan nagy arányban az ejtési szándéknak megfelelően azonosított /a:/-ra az „Álszók” kondícióban (és főként csak 494 Hz-en) több /ɒ/ válasz érkezett, mint a „Szavak” kondícióban.

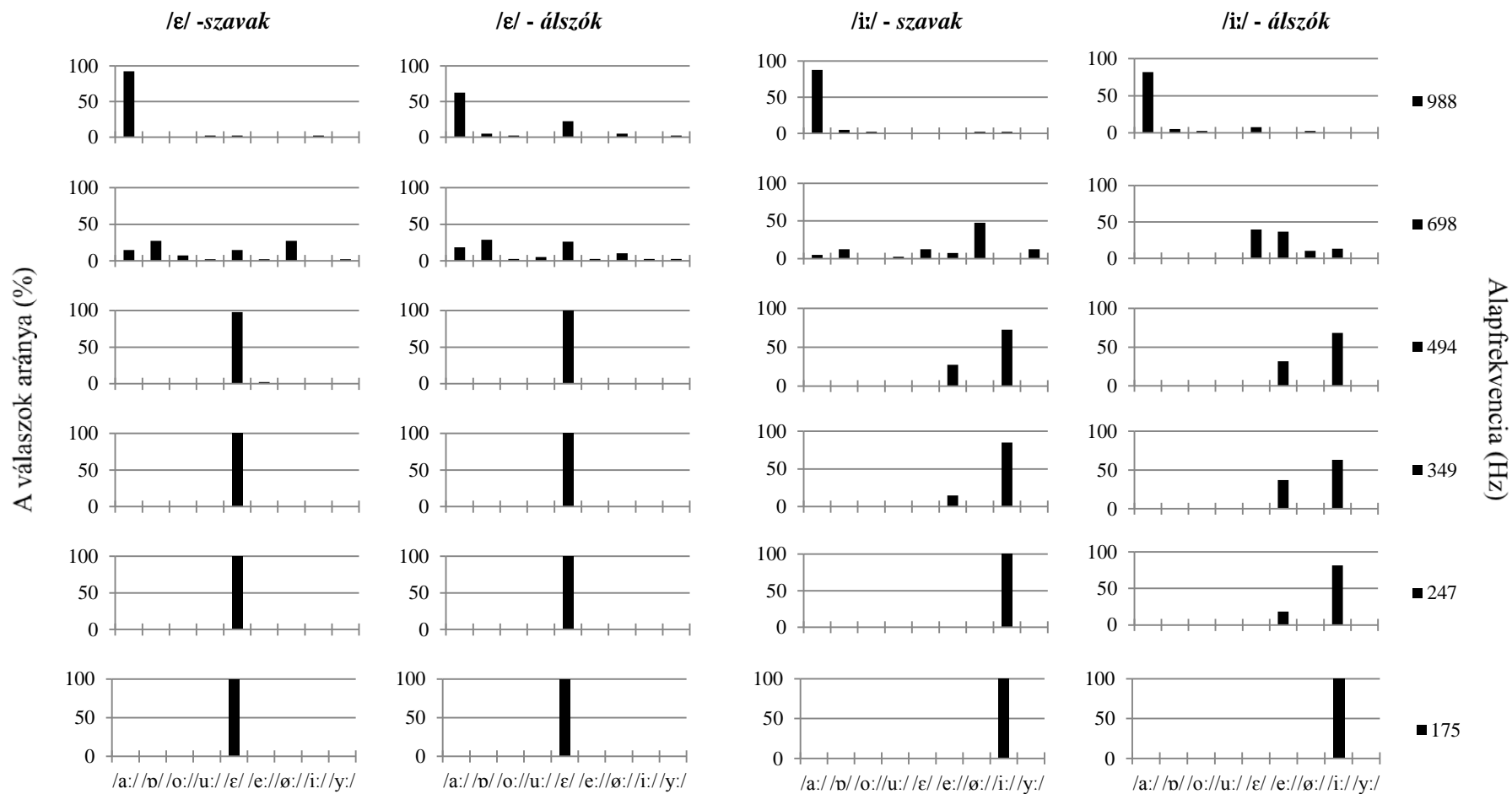
Az általános azonosítási tendenciákat összefoglaló 22. ábra alapján (mely 698 Hz alaphérfekven-  
ciától hirtelen esést mutatott) az összes választ bemutató 26. ábra figyelembevételével megállapítható, hogy a zenei *f* hangon, azaz 698 Hz alaphérfekven-  
cián a legnagyobb nyíltsági fokú /ɒ a/ magánhangzók kivételével minden vizsgált magánhangzóra közel a véletlennek megfelelő száza-  
lékban érkeztek válaszok kondíciótól függetlenül (bár sok esetben az /ɒ/ és /a/ válaszok 30% körüli dominanciájával). A zenei *h* hang, azaz 988 Hz alaphérfekven-  
cián azonban nagyon erős tendenciaként mutatkozott meg az /ɒ/ és /a/ válaszok megnövekedett aránya. A 26. ábrán látható adatok ezen túlmenően azt is megvilágítják, hogy a „Szavak” és „Álszók” kondíciók közti átlago-  
san legnagyobb, azaz 10%-nyi különbség (melyet 494 Hz alaphérfekven-  
cián tapasztaltunk) legnagyobb részt az /y:/-re kapott /ø:/ válaszoknak (rendre 11% és 19%), kisebb részben az /a:/-ra ka-  
pott /ɒ/ válaszoknak (rendre 24% és 29%), legkisebb részben pedig az /i:/-re kapott /e:/ válaszok-  
nak (rendre 17% és 18%) köszönhető. A két kondíció közti igen kis különbséget tehát összesen három magánhangzó észlelési tendenciáinak még apróbb eltérései okozzák. Ezek közül ráadásul a kondíciók hatása csak az utóbbi esetben játszhatott szerepet, hiszen az (az /e:/-vel szemben) mind az /ø:/ mind pedig az /ɒ/ legitim válasz volt mindkét kondícióban.

A 27. ábra az „énekes” csoportban kapott válaszokat összesíti alaphangfrekvenciák, magánhangzók és kondíciók szerinti bontásban. Ebben a csoportban nagy vonalakban az előzőekben bemutatottakhoz hasonló tendenciákat láttunk. Az „énekesek” egy, a kísérletet követő jellemző megjegyzése szerint az /i:/ és /e:/ ejtése énekeseknél gyakran inkább /y:/ és /ø:/-szerű realizációkat jelent (mely „csalásra” explicit énektanári instrukciók vezették korábban az énekeseket), ezért az elemzés legelején felmerült, hogy az „énekesek” esetében ezeknél a hangoknál a „nem énekeseknél” tapasztalhatóktól eltérő tendenciákat várhatunk. A megjegyzésből egyrészt következtethetnénk az /y:/ és /ø:/ válaszok preferenciájára, másrészt pedig akár éppen az /i:/ és /e:/ válaszok preferenciájára is. A 27. ábrán bemutatott adatok tanúsága szerint azonban egyik tendencia sem volt jellemző jobban az „énekesekre”, mint a „nem énekesekre”. Ugyanúgy, ahogy a „nem énekeseknél” láttuk, az „énekeseknél” is jellemzően az /ø:/-re adott válaszként láttunk nagyobb számban /y/-t (a „Szavakban” és az „Álszókban” rendre 62 és 45%), míg az /ø:/ is elsősorban az /y/-re adott válaszként jelent meg leginkább 494 Hz alaphangfrekvencián. A „nem énekes” csoporttól eltérően az /i:/-re adott /e:/ válaszok aránya a „Szavak” kondícióban is magas (mindkét kondícióban 25% körüli), azaz ők nagyobb számban választották az illegitim alakot, mint a „nem énekesek”. Ugyanakkor az is igaz, hogy 698 Hz alaphangfrekvencián a „Szavak” kondícióban lecsökkent az /e:/ válaszok aránya, és megnövekedett az /ø:/ válaszok aránya (48%), míg az „Álszók” kondícióban továbbra is az /ε/ és /e:/ dominált (rendre 39% és 37%). A „nem énekeseknél” megfigyelhettük, hogy a zenei *f''* hangon, azaz 698 Hz alaphangfrekvencián az /ɒ/ és /a:/ hangzók kivételével mind a hat vizsgált hangra nagyjából a véletlenhez közelítő arányban (de tendenciózusan az /ɒ/ és /a:/ nagyobb arányával) jelentek meg válaszok. Ehhez képest eltérést az „énekeseknél” csak az imént említett, /i:/-re kapott nagyarányú /ø:/, /ε/ és /e:/ válaszok jelentettek. Ugyanakkor ebben a csoportban is elmondható, hogy a zenei *h''* hangon, azaz 988 Hz alaphangfrekvencián egyértelműen az /ɒ/ és /a:/ hangokra „tévesztések” domináltak.

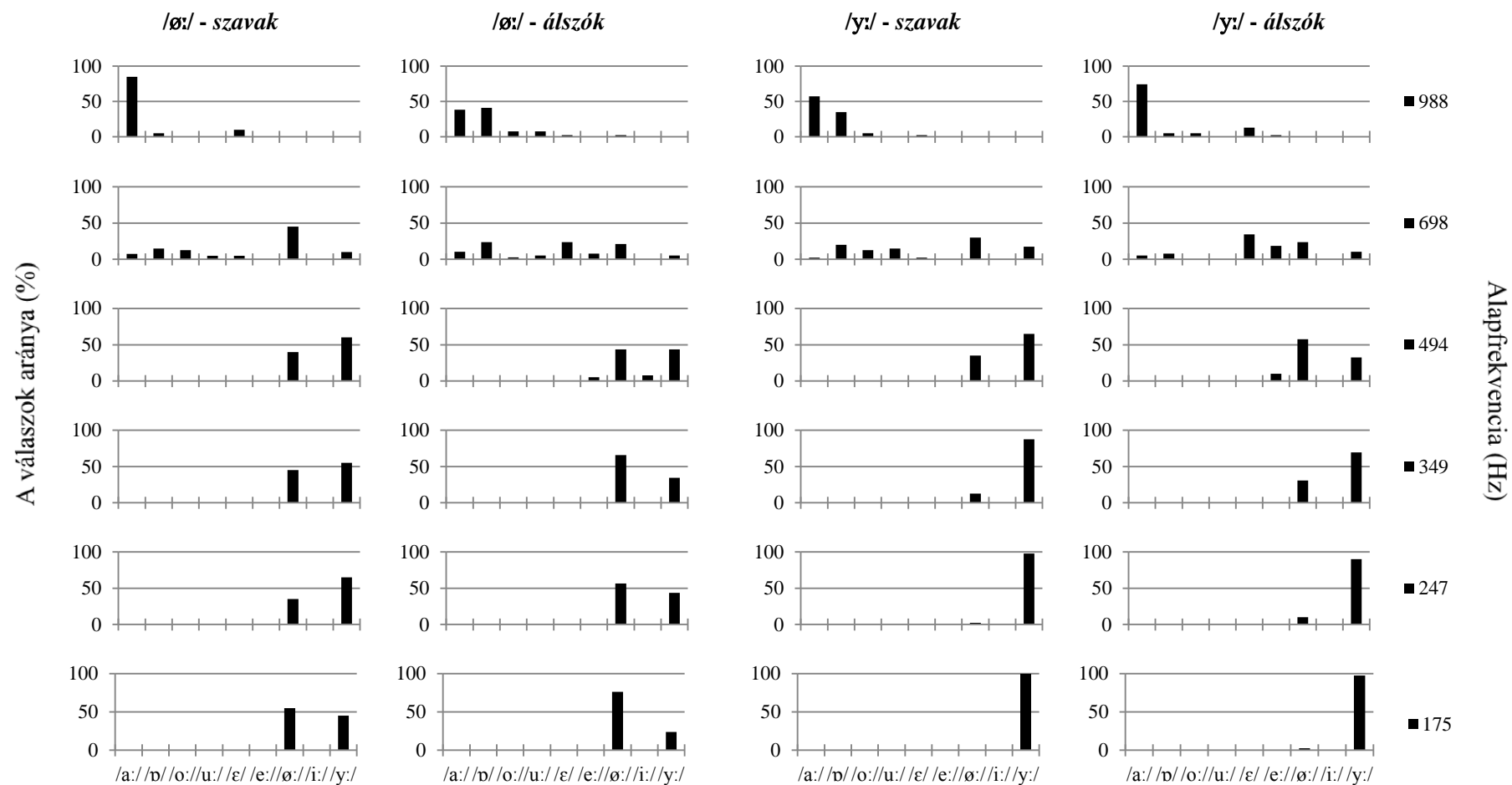


27. a ábra: Az /ɒ a:/ magánhangzóra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében az „énekes” adatközlőknél





27. b ábra: Az /ε i:/ magánhangzóra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében az „énekes” adatközlőknél



27. ábra: Az /ø: y:/ magánhangzóra adott válaszok (százalékos arányban kifejezett) alakulása a két kondícióban az f0 függvényében az „énekes” adatközlőknél

### 2.4.3. Következtetések

A kísérletben az /ɒ a: ε i: ø: y:/ magyar magánhangzók azonosítását vizsgáltuk értelmes magyar szavakban és álszókban, különös tekintettel a magas alaphangfrekvenciákra, „laikus beszélők” és énekesek csoportjában. Smith és Scott (1980) értelmes szavakban és izoláltan ejtett magánhangzók azonosítását vizsgáló kutatásából, valamint saját, a 2.3. *alfejezet*ben bemutatott értelmetlen szavakban és izoláltan ejtett magánhangzók azonosítását elemző vizsgálatunk eredményeiből azt a feltételezést fogalmaztuk meg, hogy nem hatékonyabb az ejtési szándéknak megfelelő magánhangzó-azonosítás az értelmes szavakban akkor, ha az értelmes és értelmetlen szavakban (azaz álszókban) ejtett magánhangzók azonosítását nem jelentősen eltérő (viszont akár nem is szükség-szerűen ugyanannyi), nagyszámú válaszlehetőséget megengedő feladatokban vetjük össze. A kísérletünk másik célja intuitív feltételezéseinkre és részben a beszédészlelés motoros elméletére (Lieberman–Mattingly 1985) is támaszkodva az énekes és nem énekes hallgatók teljesítményének összehasonlítása volt a magánhangzó-észlelés tekintetében. Az e célnak megfelelő második feltételezésünk szerint az énekesek nagyobb arányban képesek az ejtési szándéknak megfelelő hangzóazonosításra, mint a nem énekes hallgatók.

A kísérlet eredményei az első feltételezést alátámasztották, míg a másodikat cáfolták. Azt találtuk ugyanis, hogy sem a két vizsgált kondíció (értelmes szavak és álszók), sem pedig a két adatközlőcsoport (a nem énekesek és az énekesek) között nem mutathatók ki jelentős eltérések. Más szóval a bottom-up percepciós folyamatok aktivációja (azaz az értelmes, létező szó mint kontextus) nem segítette jobban az észlelőket a magánhangzó ejtés szerinti azonosításában akkor, ha az értelmes szavak az értelmetlen szavakhoz közel azonos számú (tehát még csak nem is teljesen egyező valószínűségű) választ engedtek meg. Megállapítottuk, hogy a legnagyobb (10%) különbség a két kondíció között (amit ráadásul kizárólag a nem énekes adatközlőknél tapasztaltunk) az /y:/-re kapott /ø:/ válaszok, valamint kisebb részben az /a:/-ra kapott /ɒ/ válaszok és az /i:/-re kapott /e:/ válaszok kondícióként eltérő számának volt köszönhető. Ezek az eltérések azonban külön-külön igen kicsinek bizonyultak, a fent említett, a kondíciók közti 10%-os különbséget csak az összesítésben eredményezték. Az első hibatípus feltehetően (legalábbis részben) az /ø:/ egyébként is megmutatkozott bizonytalan percepciós minőségével hozható összefüggésbe (amit

több ponton, így a beszéd módra kapott eredményekben is tapasztaltunk a teszt során). Többek közt a beszéd módban talált alacsonyabb azonosítási arányokból kiindulva feltételezzük, hogy az énekes adatközlő /ø:/ beszédhangjának realizációi az adatközlők perceptuális /ø:/ kategóriájának kevésbé prototipikus tagjai voltak, mely feltételezést az adatközlők a minden esetben első, értelmes szavas tesztben mutatott, az /ø:/ az ejtési szándéknak megfelelő azonosítását érintő alacsonyabb arányú teljesítménye is mutat. Következésképpen pedig – értelmezésünk szerint – az /ø:/ magasabb arányú azonosítása az értelmetlen szavakat tartalmazó második tesztben egyrészt familiarizációs hatás (az adott hangzó-megvalósuláshoz történt adaptáció), másrészt pedig hiperkorrekció eredménye lehet. Hiperkorrekción itt a következőt értjük. Mivel könnyen lehet, hogy pusztán azért, mert az első részben az adatközlők azt tapasztalták, hogy kevés egyértelműen /ø:/-szerű hangot (viszont sok az /ø:/ és /y:/ kategóriák határán álló hangot) hallottak, a második részben inkább nagyobb számban jelölték az /ø:/-t a kérdéses esetekben (tehát /y:/ vagy /ø:/ elhangzásakor). Az /p/ válaszok alacsonyabb aránya a szavak kondícióban feltehetően (legalábbis részben) magyarázható azzal, hogy az adatközlők némileg feszélyezve érezhették magukat az /p/ hangzó megjelölésével kapcsolatban a *szVr* (értelmes szavas) környezetben. Emellett az sem kizárt, hogy azt feltételezték, hogy a feladat becsapós, és ezt próbálták – ismét csak – túlkorrigálni. Végül pedig az /i:/-re kapott /e:/ válaszok arányának eltérése, mint már említettük, azért valószínű, mert az /e:/ elméletileg (az adatközlőknek adott instrukciók szerint) illegitim volt az értelmes szavas feladatrészen. Újra kiemelendő itt az az érdekes megfigyelés is, mely szerint az /e:/ illegitim státusza ellenére is gyakori válasz volt az első (értelmes szavakat tartalmazó) feladatban, ami feltehetőleg az /i:/-t érintő erős hangszínváltozásnak volt köszönhető. Mind az /y:/-re kapott /ø:/ válaszok, mind pedig az /i:/-re kapott /e:/ válaszok azokkal a korábbi vizsgálati eredményekkel csengnek egybe, melyek a produkciós és percepciós működések közti egyszerűbb, lineárisabb összefüggéseket feltételeztek, és emiatt úgy vélték, az alapfrekvencia emelésével tapasztalható produkciós nyitódást percepciós nyitódásnak kell kísérnie (l. Scotto di Carlo–Germain 1985, Hollien et al. 2000). Ugyanezen eredményeknek és feltételezéseknek megfelelő tendencia az /a:/ magas arányú azonosítása is a teljes alapfrekvencia-tartományon – hiszen az /a:/ magas első formánása miatt az alapfrekvencia emelésével akár magasabb f0 értékekig is „intakt” maradhat az F1

:  $f_0$  formánshangolástól; valamint az / $\text{p}$ / és / $\text{a}$ / válaszok megnövekedett száma 600 Hz alapfrekvencia fölött (de különösen 988 Hz alapfrekvencián).

Az alapfrekvencia emelésével általános tendenciaként megjelent a válaszok egyre nagyobb szóródása, mely a zenei  $f''$  hangon, azaz 698 Hz alapfrekvencián – a legnyíltabb / $\text{p}$   $\text{a}$ / magánhangzók kivételével – minden vizsgált magánhangzó esetében a közel a véletlennek megfelelő százalékban érkező válaszokig fokozódott (kondíciótól és résztvevőcsoporttól függetlenül). Ebből – a korábbi, 2.1., 2.2. és 2.3. *alfejezetek*ben is ismertetett vizsgálatainkkal egybehangzó módon – arra következtethetünk, hogy a magánhangzókat az alapfrekvencia emelésével egyre inkább jellemző alulspecifikáltság (azaz az artikulációs változtatások és az alul-mintavételezettség) következtében a zenei  $f''$  hang magasságának környékén az egyes hangzók minősége szinte teljesen elbizonytalanodik. Más szóval a zenei  $f''$  hang magasságának környékén kerülnek az egyes hangzók a legtávolabb az egyes prototipikus (magyar magánhangzókra jellemző) percepciók kategóriáktól. A  $h''$  zenei hang, azaz 988 Hz alapfrekvencia esetében azonban, bizonyos értelemben, ismét megerősödött a prototípushatás, ugyanis itt a válaszok között (de gyakorlatilag az ejtési szándék szerinti magánhangzó-minőségtől függetlenül) kimagaslóan megnövekedett az / $\text{p}$ / és / $\text{a}$ / válaszok aránya. 988 Hz alapfrekvencián a magánhangzó-észlelés szempontjából legfontosabbnak vélt 5000 Hz alatt bármely magánhangzó csak a 988 Hz ( $f_0$ ), 1976 Hz ( $2f_0$ ), 2964 Hz ( $3f_0$ ), 3952 Hz ( $4f_0$ ) frekvenciákon tartalmazhat intenzív összetevőket, melyek közül az első két harmonikus, az  $f_0$  és a  $2f_0$ , éppen a női ejtésben az / $\text{a}$ /  $F_1$  és  $F_2$  értékéhez közeli frekvenciaértéken helyezkedik el. A zenei  $f''$  hangon (698 Hz alapfrekvencián) erősen elbizonytalanodó, majd a zenei  $h''$  hangon (988 Hz alapfrekvencián) ismét valamivel biztosabb kategóriák azt jelzik, hogy bár az éneklésben az alapfrekvencia emelésével és a fokozatos produkciós nyílással párhuzamosan nem kategorikusan változnak az egyes magánhangzó-minőségek, de kellően magas (1000 Hz körüli) alapfrekvencián mégis tapasztalható egy kategóriális váltás. Itt ugyanis – feltehetően a felharmonikusok frekvenciaértéke miatt – a megvalósuló magánhangzók zöme a legnyíltabb / $\text{a}$ /-ként (esetleg / $\text{p}$ /-ként) azonosítható.

A nem énekes és énekes adatközlők percepcióját érintően azt láttuk, hogy az ejtési szándéknak megfelelő magánhangzó-azonosítás akkor sem volt magasabb, ha az észlelő maga is énekes, tehát ha a résztvevő már vagy az alapvetőbb akusztikai elemzésben vagy magasabb szinteken

inkorporálható artikulációs gyakorlással, illetve tudással rendelkező az éneklés artikulációs sajátosságairól. Az énekesek és nem énekesek azonosítási tendenciáiról ezen felül azt is megállapíthattuk, hogy egyéb módon (például a jellemző hibák tekintetében) sem térnek el szisztematikusan. A hipotézis cáfolatát a motoros elmélettel kapcsolatos ellenbizonyítékként értelmezhetjük akkor, ha a motoros elméletből következő feltételezésnek tekintjük azt az állítást, hogy a szenzomotoros kapcsolatok révén a produkciós gyakorlottság vagy tapasztalat percepciós előnyt is jelent(het). Eredményeink ugyanis azt jelzik, hogy a rájuk jellemző artikulációs gyakorlottság ellenére az énekesek nem látszanak percepciós előny birtokában lenni a nem énekesekhez képest. Ugyanakkor nem zárjuk ki annak a lehetőségét sem, hogy a motoros elméletből a jelen vizsgálati kérdésre más feltételezések is megfogalmazhatók volnának, melyekkel kapcsolatban eredményeink státusza nem feltétlenül egyértelmű. Ilyen feltételezés lehet az is, hogy a szenzomotoros kapcsolatok megléte és a motoros területek aktivitása a percepcióban valójában nem előnyt, hanem bizonyos tendenciák megjelenését jelenti. Ezzel kapcsolatos konklúziókra a jelen vizsgálat eredményeiből már csak sokkal óvatosabban vállalkozhatnánk, hiszen bár a két vizsgált csoport tendenciái között itt nem találtunk szisztematikus eltéréseket, ezek létét mégsem zárhatjuk ki teljesen. Ezt a kérdést egy alapvetően más jellegű (elsősorban csak az akusztikai elemzésen alapuló, értelmetlen szavakat vagy izolált ejtésű hangzókat vizsgáló) kísérletben gondoljuk vizsgálhatónak.

Első feltételezésünk teljesülését annak a felvetésnek a (némileg közvetettnek tekinthető) bizonyításaként értelmezzük, hogy a Smith és Scott (1980) vizsgálatában kapott magas eredmények, azaz a magas alaphangfrekvencián énekelt magánhangzók ejtési szándékának megfelelő azonosításának magas aránya mássalhangzó-környezetben valójában nem a kontextus hagyományos (a koartikulációs formánsátmenetek mint másodlagos akusztikai kulcsok szerepét hangsúlyozó) értelemben jelent pozitív hatást, ahogyan azt a szerzők állítják. Ennek nyomán úgy látjuk, hogy a jelen vizsgálat eredményei két fontos, a kísérlet kérdésfelvetésénél messzebbre vezető konklúziót is maguk után vonnak.

Az első konklúzió az énekelt hangzók ejtését érinti. Mivel eddig már számos artikulációs és akusztikai vizsgálat bizonyította, hogy az énekelt magánhangzók artikulációs és akusztikai célja az alaphangfrekvencia emelésével fokozatosan változik, a mássalhangzó-környezetben mint formáns-

átmenetekben kulcsolt magánhangzó-információról is csak azt feltételezhetjük, hogy erről a megváltozott artikulációval és akusztikai szerkezettel megvalósult magánhangzóról, nem pedig az ejteni szándékozott magánhangzóról szolgáltathat információkat. Mivel azonban általában nem (és a jelen vizsgálatban sem) tapasztalunk egyértelmű kategóriaugrásokat a percepció tendenciákban, továbbá a szonoráns formánsátmenetekről (a magánhangzókhoz hasonlóan) feltételezhetjük az alul-mintavételezettséget is, a kontextus illetően hatásával kapcsolatban nem szűrhetünk le egyértelmű következtetéseket. Úgy véljük azonban, hogy elvetendő az a feltételezés, melyet az éneklés szakirodalma nem ritkán sugall, és amely szerint a mássalhangzó-környezet a dinamikus akusztikai kulcsok, azaz a formánsátmenetek révén segítheti, az ejteni szándékozott magánhangzók azonosítását. Mint említettük ugyanis, ezek a kulcsok inkább csak a megváltozott minőségű hangzók akusztikai kulcsolását biztosíthatnák. A mássalhangzó-környezet mint értelmes szavakba ágyazottság ezzel szemben úgy tűnik, hogy elősegítheti az ejtési szándék szerinti azonosítást, ám pusztán azért, mert az értelmes szavak száma és a bennük előforduló magánhangzók minőségi különbségei növelik a „találat” valószínűségét ahhoz a felismerési feladathoz képest, ahol az összes magyar (vagy más nyelvű) magánhangzó lehetséges válasz lenne.

A második konklúzió az eddig (a 2.3. és 2.4. *alfejezetekben*) megfogalmazott egyik javaslatunk megerősítése, mely az előzőekből is következően a mássalhangzó-környezetnek tulajdonított pozitív hatások vizsgálatában a „mássalhangzó-környezet” terminus differenciálásának szükségességére hívja fel a figyelmet (mind az éneklés mind pedig a beszéd szakirodalmában). Azt tapasztaltuk ugyanis, hogy míg a „mássalhangzó-környezet” mint **koartikulációs formánsátmenet** nem segíti a percepciót (l. a jelen értekezés 2.3. *alfejezetét* és a jelen alfejezetet Smith és Scott 1980 eredményeivel összevetve), addig azt is láttuk, hogy a „mássalhangzó-környezet” mint **értelmes szavakba ágyazottság** mégis hatást gyakorol, hiszen behatárolja, korlátozza a válaszlehetőségeket (a top-down percepció folyamatok aktivációja révén), ami kedvez az ejtési szándéknak megfelelő azonosításnak is (l. Smith és Scott 1980 eredményeit). Az alfejezet zárlataként következtetéseinket a fogalom két aspektusára külön-külön reflektálva foglaljuk össze (ezzel feloldva a vizsgálatok közti látszólagos ellentmondást).

Eddigi eredményeink, valamint Smith és Scott (1980) eredményei fényében állíthatjuk, hogy a magas alapfrekvencián énekelt magánhangzók esetében mind a magánhangzókat, mind pedig a

szomszédos mássalhangzók és a magánhangzó közti koartikulációs formánsátmeneteket alulspecifikáltság (azaz spektrális alul-mintavételezettség, illetve megváltozott artikuláció és akusztikum) jellemzi. Ebből kifolyólag a kontextus, azaz **a koartikulációs formánsátmenetek** nem támogatják (de nem is támogathatják) az ejtési szándéknak megfelelő azonosítást. Ugyanakkor azt is állíthatjuk, hogy a kontextus mint **értelmes szavakba ágyazottság** jelentős támaszt jelenthet az ejtési szándék szerinti hangzóazonosítás számára, hiszen ha a mássalhangzózváz kellően kisszámú lehetőséget enged meg, az megnöveli az egyes „megoldások”, kiegészítések valószínűségét, tehát az azonosítás „sikerességének” valószínűségét is (de az akusztikai információktól ilyen értelemben függetlenül). Következésképpen úgy látjuk, hogy a magas alapfrekvencián énekelt magánhangzók azonosítása nem az elsődleges vagy másodlagos akusztikai kulcsok, tehát nem az akusztikai feldolgozás, hanem szinte kizárólag a kontextus, valamint a fonológia függvénye: a tágabb szöveggörnyezetbe, és mondatba ágyazottságtól, valamint a fonotaktikai valószínűségek hatásaitól és a fonológiai szomszédsági hatástól (a fonológiai szomszédok, azaz a csak egy fonémában eltérő szavak számától, idegen szóval a *phonological neighborhood density* hatásaitól) függ.

## 2.5. Az énekelt magánhangzók akusztikai szerkezete néhány módszertani újítás tükrében, valamint az akusztikai és percepciós jellemzők összefüggései

A 2.1. *alfejezet*ben bemutatott vizsgálatunkban a magas alapfrekvenciájú éneklésben megvalósuló magyar magánhangzókat vizsgáltuk akusztikai és percepciós szempontok szerint. Az akusztikai elemzés során kiemeltük, hogy magas alapfrekvencián a magánhangzó-realizációk formánsai, azaz a toldalékcsőre mint üregrendszerre az ejtés közben jellemző sajátrezonanciák nem állapíthatók meg egyértelműen az akusztikai jelből. Amint azt az 1.1.1.2. *alfejezet*ben részletesen tárgyaltuk, a magas alapfrekvencián ejtett hangok esetében az alacsony alapfrekvencián megvalósított hangzókhoz képest kisebb sűrűségű a felharmonikus-szerkezet, tehát a hangzók spektrálisan alul-mintavételezetté válnak, ami az  $f_0$  emelésével egyre fokozódik. Ennek következtében a magas alapfrekvenciák esetén az akusztikai jel Fourier-transzformáción alapuló elemzése kizárólag



a spektrális energiamaximumokat teszi detektálhatóvá, nem pedig az azokat előidéző rezonanciákat. A rezonanciák frekvenciahelyeit ugyanis az egyes felharmonikusok erősítettsége mindössze csak sejteni engedi, azon keresztül, hogy az erősített felharmonikusok valószínűsíthetően beleesnek valamely rezonancia sáv szélességébe. Mindennek következtében az akusztikai jel elemzése magas alapprofrekvenciákon már nem enged megbízhatóan következtetni az ejtést jellemző artikulációs működésekre és akusztikai stratégiákra. Éppen ezért a szopránénekülésben a magas alapprofrekvencián feltételezett  $F1 : f_0$  hangolás pusztán az akusztikai kimeneti jel (azaz az énekhang) vizsgálatán keresztül nem bizonyítható, annak ellenőrzéséhez olyan alternatív mérési eljárás szükséges, mely lehetővé teszi a toldalékcső sajátrezonanciáinak az alapprofrekvenciától független mérését.

A jelen vizsgálat elsődleges célja, hogy megállapítsa azokat az akusztikai változásokat, melyek a magyar magánhangzók formánsszerkezetét jellemzik az alapprofrekvencia emelésével, és ily módon bizonyítékot szerezzünk a magas alapprofrekvenciákon ( $f_0 > F1_{\text{beszéd}}$  helyzetben) feltételezett  $F1 : f_0$  hangolási stratégiáról. Ehhez a vizsgálatban két olyan módszertani újítást alkalmazunk, melyek korábbi vizsgálatok tanúsága szerint alkalmasak lehetnek arra, hogy segítségükkel még magas alapprofrekvencián is megbízhatóan detektáljuk a toldalékcsőre jellemző sajátfrekvenciákat. Az egyik újítás, hogy az audiojel mellett a hangszalagok működését is regisztráljuk elektrolottográf segítségével, és a formánsfrekvenciákat az akusztikai jel inverz szűrésével állapítjuk meg olyan módon, hogy a szűrésben az elektrolottogramban rejlő információkat is hasznosítjuk (vö. Hertegård és Gauffin 1993). A másik újítás pedig, hogy rögzítünk és elemzünk egy, a toldalékcső külső gerjesztése mellett felvett audiojelet is (vö. Sundberg 1975), így erősítve meg (és egészítve ki) az inverz szűrés során nyert akusztikai adatokat. Az inverz szűrést és az elektrolottográfiát, valamint ezek felhasználását a formánsmérésekben a következő két (2.5.1. és 2.5.2.) alfejezetben mutatjuk be. A külső gerjesztés használatát és a „toldalékcső újramintavételezésének” módszerét pedig a 2.5.4. alfejezetben ismertetjük.

A jelen vizsgálat másodlagos célja, hogy az akusztikai adatokat összefüggésbe hozzuk az e magánhangzók azonosítását jellemző tendenciákkal. Ehhez az akusztikai elemzéseken kívül egy, az előzőekhez nagyban hasonló észlelési tesztet is végzünk. Mivel azt reméljük, hogy akusztikai elemzéseink során minden eddiginél pontosabb képet kapunk az egyes magánhangzók ejtésekor jellemző formánsfrekvenciákról, így az is lehetségessé válik, hogy meghatározzuk, mely

felharmonikusok válnak erősítetté, és mely rezonanciák által. Ezt az információt pedig megkísér-  
reljük felhasználni az észlelési eredmények értelmezésében is, mégpedig Traunmüller (1981)  
modelljén keresztül, mely a magánhangzók észlelt zártságát elsősorban az F1 és az f0  
(tonotópiás) távolságából eredezteti. Az elméletről átfogóan már az *1.1.2.2. fejezetben* írtunk, a  
hozzá kapcsolódó kísérletet és az elméletnek a jelen kérdéskörre vonatkoztathatóságát pedig az itt  
következő *2.5.3. fejezetben* ismertetjük.

### **2.5.1. Az elektroglottográfia és az elektroglottogram**

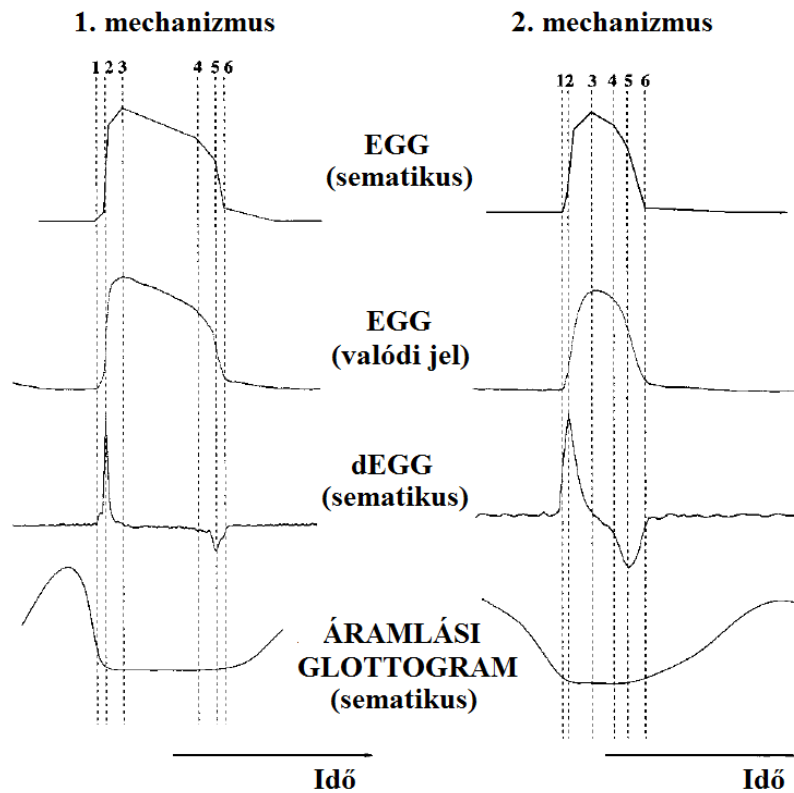
Az itt következőkben a vizsgálatban használt egyik technológiát, az elektroglottográfiát mutat-  
juk be olyan szempontok alapján, melyek megvilágítják, miért és hogyan használható ez az eljá-  
rás a formánsmeghatározásban.

Az elektroglottográf olyan (az egészségre teljességgel ártalmatlan) eszköz, mely segítségével  
detektálható a hangszalagok zárt állása a zöngképzés során. Az elektroglottográf működési elve,  
hogy két elektróda segítségével kisfeszültségű, magas frekvenciájú elektromos áramot vezet át a  
pajzsporc egyik oldaláról a másikra a hangszalagok magasságában, és az átjutó jel feszültségvál-  
tozásait rögzíti. Ezt a jelet hívják elektroglottogramnak (EGG)<sup>21</sup>. Mivel a hangszalagok közelítése  
és távolítása miatt az átjutó áram mennyisége a zöngeciklus lefolyásával változik, az így kapott, a  
feszültségváltozást az idő függvényében rögzítő hullámforma, azaz az elektroglottogram a hang-  
rés nyitódásáról és záródásáról, azaz a hangszalagok mozgásáról informál. Ugyanezen okból ki-  
folyólag azonban az is igaz, hogy az EGG csak akkor igazán pontos, amikor a hangszalagok zárt  
állásban vannak (vagy éppen záródnak): a számos (az EGG használatát felderíteni szándékozó)  
exploratív vizsgálat (csak) azt bizonyította egyértelműen, hogy minél nagyobb a hangszalagok  
érintkezésének felülete, annál nagyobb az átmenő feszültség (Henrich et al. 2004, Herbst et al.  
2010). Az EGG-nek, valamint annak első deriváltjának (dEGG) egy ciklusa a zöngeciklus szaka-  
szainak megfelelően több elkülönülő részre osztható, ezt szemléltetjük a 28. ábrán (110 Hz alap-  
frekvencián, egy férfi ejtésében, modális és falzett zöngképzésben, melyet a nemzetközi szak-  
irodalomban bevett terminológia alapján az ábra forrását képező tanulmányban 1. és 2. fonációs

---

<sup>21</sup> Az EGG rövidítés a szakirodalomban mind az elektroglottográfiát mint eljárást, mind pedig az így rögzített hul-  
lámformát, azaz az elektroglottogramot is jelöli.

mechanizmusnak neveznek, vö. még pl. Titze 2004). Az EGG hullámforma szakaszokra osztását Henrich és munkatársainak (2004) összegző elemzése alapján ismertetjük. A zöngeciklus szakaszainak megfelelően az ahhoz rendelt EGG-ciklus négy részre osztható.



28. ábra: Az EGG és dEGG jelek illusztrációja a modális (1. mechanizmus) és falzett (2. mechanizmus) zöngéképzésben, valamint az EGG és a dEGG szinkronizációja az áramlási glottogrammal a zöngéképzés egy ciklusában: (1)–(3): a hangszalagok összezáródása; (3)–(4): zárt állás; (4)–(6): a hangszalagok nyitódása; (6)–(1): nyitott állás (Henrich et al. 2004 alapján)

(1)–(3): a hangszalagok összezáródása: a záródás a hangszalagok alsó éleinél kezdődik ((1)-től (2)-ig), majd továbbterjed a hangszalagok felső éléhez ((2)-től (3)-ig). Mivel a záródás általában gyorsabb, mint a nyitódás, ez a szakasz nagy (pozitív) meredekségű. A maximális meredekséget (*maximum slope*), azaz azt a pontot, amikor az elektromos feszültség a leggyorsabban változik a (2) pont jelöli. A későbbiekben még látni fogjuk, hogy ez a pont kitüntetett jelentőségű, így az

elemzésben gyakran a megállapítására kifejezetten alkalmas első deriváltat (dEGG) használjuk. (Mivel az első deriváltban a hirtelen változásokat a meredekségtől függően kisebb vagy nagyobb kiugrásokként látjuk, ez a pont a dEGG-ben nagy intenzitású kiugrásként jelenik meg, ehhez l. még Herbst et al. 2010.)<sup>22</sup>

(3)–(4): zárt állás: a hangszalagok teljesen (vagy majdnem teljesen) összeérnek, így a glottiszon átáramló levegő mennyisége is nullára csökken (az itt tapasztalható csekély feszültségnövekedést vagy -csökkenést egyes kutatók az egymásnak csapódó rugalmas felületek dinamikájával magyarázzák).

(4)–(6): a hangszalagok nyitódása: a hangszalagok alsó éleinek fokozatos szétválását ((4)-tól (5)-ig) a felső élek szétválása követi ((5)-tól (6)-ig). A maximális (negatív) meredekséget itt az (5)-ös pont jelöli, melyet (éppen ezért) a dEGG-n negatív csúcsként látunk. Mivel a hangszalagok szétválása lassabb, mint az összezáródás, ezért itt a maximális meredekség is kisebb, mint a ciklus másik felében, és a dEGG-n csak jóval kisebb csúcs jelenik meg, mint a záródási szakasz megfelelő pontján.

(6)–(1): nyitott állás: a hangszalagok nyitott állásban állnak. Az itt mért feszültség relatíve alacsony, és csak kismértékű ingadozást mutat (Henrich et al. 2004).

A dEGG-n (2)-es számmal jelzett „záródási csúcs”, valamint az (5)-ös számmal jelzett „nyitódási csúcs” különös jelentőséggel bír. Ezekről ugyanis (párhuzamos videokimográfias és inverz szűrésen alapuló összehasonlító elemzéseken keresztül) kimutatták, hogy valóban a hangszalagok találkozási és szétválási pontjára utalnak (Henrich et al. 2004). Mivel azonban a záródási szakasz gyorsabb, így a „záródási csúcs” valamivel egyértelműbb és megbízhatóbb referenciapontnak tekinthető.

---

<sup>22</sup> Megemlítendő, hogy bár a záródási pont prototipikusan egyetlen hirtelen kiugrás, egyes beszélőknél kétszúcsú is lehet. Ezt a más képalkotó eljárással készített összehasonlítások alapján „cipzárszerű” záródás eredményének tartják, azaz a hangszalagok felületével párhuzamos előlről hátrafelé terjedő (*anterior-to-posterior*), vagy hátulról előre felé terjedő (*posterior-to-anterior*) záródás lenyomatának, de részletes vizsgálatok híján a jelenségre valójában többféle magyarázatot is elképzelhetők tartanak (erről részletesebben l. Henrich et al. 2004, Herbst et al. 2010).

### 2.5.2. Az inverz szűrés, az áramlási glottogram, valamint az elektroglottográfia szerepe az inverz szűrésben

Az inverz szűrés a beszédképzés hagyományos, a forrást és szűrőt élesen elválasztó modelljéből indul ki, melynek alapfeltételezése a forrás és szűrő lineáris összefüggése is. Az eljárás elsődleges célja, hogy megbecsülje a glottiszon a beszédképzés (vagy más hangadás, illetve a légzés) alatt áthaladó levegő mennyiségét, valamint a levegő térfogatsebességét az idő függvényében. Ez a függvény az áramlási glottogram (*glottal flow, airflow*), mely tehát a hangréseken átáramló levegő mennyiségét ábrázolja miliméter/másodperc [mm/s] mennyiségként az idő függvényében. Az inverz szűrés során a toldalékcső átviteli karakterisztikáját leképező szűrősor inverzével, azaz az inverz szűrősorral kompenzáljuk a toldalékcső rezonátortevékenységét, így „állítva helyre” a gégénél létrejövő áramlási glottogramot<sup>23</sup> (Hertegård–Gauffin 1993). Bár az eljárás elsődleges célja nem ez, de működési elvéből következően a rezonanciahelyek detektálására is használható.

Az inverz szűrés a jelen alfejezetben bemutatott vizsgálatban is a formánsmeghatározás egyik módszere, melyhez EGG-t is használunk (Hertegård–Gauffin 1993 elemzéséhez hasonlóan). Vizsgálatunk módszertana szempontjából fontos megfigyelnünk a 28. ábrán az EGG és dEGG, valamint az áramlási glottogram összefüggéseit és megfeleltethetőségét. Ahogy már említettük, az EGG-ben kifejezett jelentőséggel bír a hangszalagok találkozásának pontja, a dEGG-ben megfigyelhető „záródási csúcs”. A transzglottális levegőáram szempontjából ugyanis a záródási csúcs éppen annak a szakasznak a kezdetét jelöli, ahol a levegőáram mennyisége hirtelen nagyon lecsökken – tehát mindkét hullámformán egyértelműen beazonosítható a zárszakasz kezdete. Bár megfigyelhető az is, hogy a nyitódási csúccsal egy időben a levegőáram mennyisége is lassú növekedésnek indul, de jól látható, hogy az áramlási glottogram e szakaszának meredeksége jóval kisebb a záródási szakaszban tapasztalhatónál. Ennek oka, hogy míg a záródás során a levegő útját hirtelen összezáródással elálló hangszalagok valóban képesek hirtelen csökkenést okozni a glottiszon átáramló levegő mennyiségében, addig az újra kinyíló hangréseken átáramló levegő se-

---

<sup>23</sup> Fontos itt megemlíteni, hogy a forrást és a szűrőt csak a modell értelmezésében, az elemzés megkönnyítése érdekében tekintjük függetlennek. Ezek valójában szoros kölcsönös függésben állnak egymással. Ennek értelmében tehát az inverz szűrés során visszafejtett jel is csak elméleti reprezentációja lehet a hangréseken átáramló levegőnek. Igaz ugyanakkor az is, hogy sem a hangréseken átáramló levegő, sem pedig a gégénél keletkező hangnyomásingadozás, azaz a zöngé sem mérhető vagy rögzíthető közvetlenül, ilyen értelemben tehát ezekről nagyon jó közelítést kaphatunk az inverz szűrés által.

bessége a levegőoszlop nagyságától és tehetetlenségétől függ – így ezt a szakaszt jóval lassabban elinduló mozgás és kisebb áramlási sebesség jellemzi. Az inverz szűrés lényege, hogy a toldalécső rezonátortevékenységet kompenzálva helyreállítsuk a gégénél létrejövő áramlási glottogram hullámformáját, ami elsősorban a fodrozódásmentes zárszakasz-hullámforma helyreállítását jelenti. A dEGG pedig a záródási csúcs révén jelzi azt a pontot, ahol a zárszakasz kezdődik, és ahová az áramlási glottogram zárszakaszt jelző hajlatának esnie kell (ez tehát a (2)-es pont). Ezért az EGG és különösen a dEGG nagy segítség a manuális inverz szűrés során, különösen akkor, ha magasabb alaphfrekvenciákon ejtett hangokat elemzünk falzett regiszterben (vagy más szóval a zöngképzés 2. mechanizmusában), melyek áramlási glottogramjának hullámformája a modális, alacsony frekvenciájú zöngé hullámformájánál rövidebb hullámhosszal és rövidebb zárszakasszal jellemezhető (l. pl. Henrich et al. 2004, valamint a 28. ábrát).

Az inverz szűrést az énekelt magánhangzók formánshelyeinek meghatározására eddig Hertegård és Gauffin (1993), valamint Sundberg és munkatársai (2011, 2013) használták sikerrel – utóbbiak férfi hangosztályok alacsonyabb alaphfrekvenciájú énekprodukciónak. Hertegård és Gauffin (1993) tanulmányának fő konklúziója, hogy az inverz szűrés EGG alkalmazásával megfelelő módszer lehet még a magas alaphfrekvencián ejtett magánhangzók formánsainak detektálásában is.

Mivel a korábbi eredmények alapján a magas alaphfrekvencián énekelt beszédhangokat érintő legfontosabb akusztikai változás az F1 frekvenciájának (az f0 emelését kísérő) emelése, az észlelési vizsgálatban is a magánhangzók (az ejtési szándéknak megfelelően vagy attól eltérően azonosított) nyíltsági fokára koncentráltunk. Ezzel egyetértésben észlelési adatainkat a zártságot az F1 és f0 tonotópiás távolságából levezető elmélet kereteiben (Traunmüller 1981, l. következő alfejezet), az akusztikai adatainkat is felhasználva értelmeztük. Mivel tudjuk, hogy a hallási észlelési folyamatok frekvenciaérzékenyek, az akusztikai adatokat érzetileg (logaritmikusan) transzformáltuk, továbbá a frekvenciaérzékenység miatt is létrejövő a magas alaphfrekvencián ejtett magánhangzókra jellemző alacsony akusztikai felbontás miatt azt feltételeztük, hogy nem lehet egyszerű összefüggéseket vonni a „nyers” formánsfrekvenciák és az észlelet között. A jelen vizsgálat elemzési kísérlete ilyen módon az énekelt beszédhangok kutatásában újdonságnak számít, segítségével azonban módunk nyílik magyarázatot találni a jelen értekezés korábbi alfejezeteiben is-

mertetett vizsgálatok egyes, akkor magyarázat nélkül maradt észlelési tendenciáira – a zárt ejtésű hangzók jó arányú azonosítására magas alapfrekvencián, valamint a tévesztési tendenciákban az egyes hangzók zárt ejtésű hangzókként való azonosítására.

### **2.5.3. Traunmüller modellje és a modell predikciói az énekelt magánhangzók azonosítására**

A szakirodalomban olvasott, mi több, azt meghatározónak látszó feltételezés szerint az alapfrekvencia emelésével az F1: f0 hangolás következtében egyre nyíltabban ejtett hangzókat olyan beszédhangokkal tévesztik össze a hangokat azonosító hallgatók, melyek a beszédben nyíltabb ejtésűek, illetve magasabb F1 formánssal képzettek, mint az ejteni szándékozott beszédhang (l. pl. Scotto di Carlo–Germain 1985, Hollien et al. 2000). Ugyanakkor azt is láttuk, hogy e feltételezések mögött nem áll megbízható és erős empirikus bizonyíték. Ráadásul egyes saját eredményeink (l. a jelen értekezés 2.1. és 2.2. *alfejezetét*) e feltételezésekkel gyökeresen ellentétesnek látszottak lenni, hiszen saját vizsgálataink a zárt ejtésű magánhangzók jó arányú azonosítását mutatták ki közepesen magas alapfrekvenciákon, valamint azt a tendenciát, hogy a magas alapfrekvenciákon ejtett hangzókat az észlelők nagy számban azonosítják zárt hangzókként.

Mivel korábban nem álltak rendelkezésünkre megbízható rezonanciaadatok az egyes hangzók ejtésekor jellemző formánsokról, nem vizsgálhattuk adatainkat például azon elmélet kereteiben, mely a zártságészleletet az F1 és f0 tonotópiás távolságából (az f0 és 2f0 távolságával, tehát az alapfrekvenciával összefüggésben) vezeti le. A jelen kísérletben azonban az énekelt magánhangzókról először formánsadatokat nyertünk, így ez alkalommal lehetőségünk volt a percepciósi adatokat az (érzetileg transzformált) rezonanciák fényében értelmezni.

Traunmüller (1981) vizsgálatait olyan osztrák dialektusok beszélőivel készítette, mely dialektusokban ötféle fonológiai kontrasztív nyíltsági fok különül el. Az első fok a legzártabb, míg az ötödik fok a legnyitottabb hangzókat jelöli. Ezek rendre a következők: 1: /i u y/; 2: /e o ø/; 3: /ɛ ɔ œ/; 4: /æ ɒ œ/; 5: /a/. Traunmüller ezen hangzók segítségével vizsgálta azt a kérdést, hogy milyen szerepet játszik az f0 és az F1 tonotópiás, azaz érzeti távolsága a magánhangzók zártságának észleletében.

A modell a már Chistovich és Lublinskaja (1979) által kimutatott 3 Bark másodlagos kritikus sávon alapszik (elsődleges kritikus sávnak az 1 Barkot tekintjük, mely azt a frekvenciafüggő frekvenciasávot jelöli, ami – az alaphártya felépítéséből következően – a fül frekvenciafeldolgozásának legkisebb egysége). A másodlagos, 3 Barkos kritikus sáv integrációs határként értelmezhető: azok az akusztikai prominenciák, melyek 3 Bark észleleti távolságnál kisebb távolságra helyezkednek el egymástól, az észleletben egyetlen, frekvenciájában a két prominencia frekvenciaértéke közötti (azok amplitúdójával súlyozott) „gravitációs középpontba”, azaz egyetlen prominenciába integrálódnak.

Traunmüller (1981) modellje szerint abban az esetben, ha két harmonikus (tehát például az  $f_0$  és  $2f_0$ ) között, melyek egy rezonancia (pl.  $F_1$ ) közelébe esnek, kisebb a távolság, mint 3 Bark, a két harmonikus mint akusztikai prominencia integrálódik, mégpedig a közeli (a példában az  $F_1$ ) rezonanciának az értékéhez közeli prominenciában. Mivel 370 Hz alatt az  $f_0$  és a  $2f_0$  távolsága kisebb, mint 3 Bark, így itt az  $f_0$  és a  $2f_0$  minden esetben (magánhangzó-minőségtől függetlenül) egy érzeti csúcsot eredményez. Mivel pedig az  $f_0$  és az  $F_1$  az első, második és harmadik nyíltsági fokú magánhangzó esetében szintén 3 Barkon belül esik, így ezen hangzók esetében az  $f_0$  és  $F_1$  tartományán mindösszesen ez az egy érzeti prominencia jelenik meg, ami tehát (megközelítőleg) az  $F_1$  értékét veszi fel. 370 Hz alatt a Traunmüller kísérletében megkülönböztetett első három nyíltsági fokú magánhangzó úgy különül el egymástól, hogy az egyes és kettes, valamint a kettes és hármas nyíltsági fokú magánhangzók közti különbséget az  $F_1$  és  $f_0$  távolsága – vagy másként a fentiek értelmében az  $F_1$  értéke – adja. Ha az  $F_1$  és  $f_0$  különbsége kisebb vagy egyenlő, mint 1,2–1,5 Bark, a hangot egyes nyíltsági fokúnak észleljük (/u i y/). Ha az  $F_1$  és  $f_0$  különbsége nagyobb, mint 1,2–1,5 Bark, de nem nagyobb, mint 2,5 Bark, a hangzót a második nyíltsági fokba tartozónak észleljük (/e o ø/). Ha pedig az  $F_1$  és  $f_0$  különbsége nagyobb, mint 2,5 Bark, de kisebb, mint 3–3,5 Bark, a hangzót a harmadik nyíltsági fokba tartozónak észleljük (/ε ɔ œ/). Mivel az  $F_1$  és  $f_0$  e hangok esetében 370 Hz alatt az  $F_1$  értékének megfelelő frekvenciájú érzeti prominenciába integrálódik, mindezt úgy is megfogalmazhatnánk, hogy a zártság észleletét az első, második és harmadik nyíltsági fokú beszédhangok esetében az  $F_1$  értéke befolyásolja a beszédben.



A negyedik és ötödik nyíltsági fokú magánhangzók esetében azonban már más a helyzet, hiszen ezekben az  $f_0$  és az  $F_1$  értéke közti különbség nagyobb, mint 3 Bark. A negyedik és ötödik nyíltsági fokú hangokban 370 Hz alapfrekvencia alatt tehát az  $f_0$  és  $F_1$  nem egy, hanem két érzeti prominencia jelenik meg az  $F_1$  és  $f_0$  frekvenciatartományán. Az elmélet szerint a negyedik nyíltsági fokú hangokat az  $F_1$  és  $f_0$  között lévő 3 Barknál nagyobb távolság definiálja az észlelés számára: ha az  $F_1$  és  $f_0$  közti különbség nagyobb, mint 3 Bark, de kisebb, mint 5,5–6 Bark, a hangzót a negyedik nyíltsági fokú magánhangzónak észleljük (/æ ɒ œ/), míg ha az  $F_1$  és  $f_0$  különbsége ennél is nagyobb, az észleletben a legnyíltabb /a/ jelenik meg. A 370 Hz alapfrekvencia alatt ejtett (negyedik és ötödik nyíltsági fokú) hangzók további jellemzője az is, hogy ezekben az első (az  $f_0$ -hoz eső) érzeti csúcs kevésbé prominens, mint a második ( $2f_0$ -hoz eső) csúcs, ugyanis a magas frekvenciájú  $F_1$  ez utóbbit erősíti.

Mindez azonban hangsúlyozottan csak a 370 Hz alapfrekvencia alatt ejtett hangzókra igaz. 370 Hz alapfrekvencia fölött ugyanis már az  $f_0$  és az azt követő első felharmonikus ( $2f_0$ ) közti távolság is átlépi a kritikus integrációs (3-3,5 Barkos) határt. A kritikus határ átlépése a fentebbi okfejtés értelmében azt eredményezi, hogy a 370 Hz alapfrekvencia fölött ejtett beszédhangok észleletében minden esetben két érzeti csúcs (az  $f_0$  és a  $2f_0$ ) jelenik meg a kérdéses tartományon, függetlenül az  $F_1$  értékétől, tehát még akkor is, ha az első három nyíltsági fokba tartozó hangzók-ról van szó. Könnyen belátható, hogy a modell szerint ez egyúttal azt is jelenti, hogy 370 Hz alapfrekvencia fölött a beszédhangok a 370 Hz alapfrekvencia alatt ejtett negyedik nyíltsági fokú hangzókra kezdenek hasonlítani, hiszen a két érzeti prominencia helye a 370 Hz alapfrekvenciánál magasabb alaphangon ejtett hangokban a 370 Hz alapfrekvencia alatt ejtett negyedik nyíltsági fokú beszédhangokéval (kvázi) megegyező.

A modell tehát azt jósolja, hogy a különböző magánhangzók megvalósulásai 370 Hz alapfrekvencia fölött – a hangzók észleleti prominenciáitól vagy valós  $F_1$  értékétől függetlenül – majdnem minden hangzóra ugyanazt az „észleleti eredményt” adják. A „majdnem” azonban nagyon fontos, hiszen ha ténylegesen és pontosan a modell predikcióit követve próbáljuk értelmezni a 370 Hz felett ejtett beszédhangok azonosítását, akkor azt találjuk, hogy a 370 Hz alapfrekvencia alatt ejtett negyedik nyíltsági fokú beszédhangokhoz valójában csak a 370 Hz alapfrekvencia fölött ejtett második és harmadik nyíltsági fokú hangzók válnak egyértelműen hasonlatossá. A 370

Hz alaphfrekvencia fölött ejtett második és harmadik nyíltsági fokú beszédhangokban ugyanis az F1 frekvenciája magasabb, inkább a  $2f_0$ -hoz közeli, így tehát a két érzeti prominencia ( $f_0$  és  $2f_0$ ) közül ezekben a második csúcs lesz intenzívebb, prominensebb –és ez valóban a 370 Hz alaphfrekvencia alatt ejtett negyedik nyíltsági fokú hangokhoz hasonló mintázat. Az első nyíltsági fokú (/i u y/) hangzók esetében azonban más a helyzet, és ezekre a modell nem is fogalmaz meg egyértelmű jóslatot. Az első nyíltsági fokú beszédhangokban ugyanis (370 Hz fölötti alaphfrekvencián ejtve) nem a második, hanem az első, az  $f_0$  közelébe eső prominencia lesz intenzívebb, prominensebb, hiszen az F1 itt az  $f_0$ -hoz esik közel, azt erősíti. Ezt a prominenciamintázatot viszont a negyedik nyíltsági fokú hangzóknál sosem tapasztaljuk „normál”, 370 Hz alaphfrekvencia alatti ejtésben. Így tehát az első nyíltsági fokú (zárt, felső nyelvvállású) beszédhangok – a többi nyíltsági foktól eltérően – 370 Hz fölött nem alakulnak az észleletben törvényszerűen negyedik nyíltsági fokú hangzókká. De akkor vajon mégis mi történik ezekkel a zárt ejtésű hangzókkal 370 Hz alaphfrekvencia fölött az észlelésben? Erre a kérdésre Traunmüller (1981) modellje nem fogalmaz meg egyértelmű predikciókat, a kísérleti eredményei (Traunmüller 1981) alapján azonban az látszik, hogy az első nyíltsági fokú hangzók a teljes általa vizsgált alaphfrekvencia-tartományon (megközelítőleg 700 Hz-ig) felismerhetőek voltak, és elkülönültek a többi hangzótól. Ebből következően tehát, egészen érdekes módon, a tonotópiás modell szerint 370 Hz-nél magasabb alaphfrekvenciákon csak első és negyedik (illetve ötödik) nyíltsági fokú hangzókat, azaz csak a zárt, felső nyelvvállású vagy első nyíltsági fokú és a nyílt, negyedik (vagy ha az a negyediktől elkülönül, ötödik) nyíltsági fokú, alsó nyelvvállású beszédhangokat képes azonosítani a beszédpercepció – azokat viszont igen magas alaphfrekvenciáig képes elkülöníteni egymástól.

Az énekelt magánhangzók azonosításának kérdését is szem előtt tartva a fentieket a következőképpen foglalhatjuk össze. A modell szerint 370 Hz alaphfrekvencia (mely egyébként az  $f_{\#}'$  zenei hang alaphfrekvenciája) alatt a zártság észlelete az F1 és  $f_0$  érzeti távolságának, valamint az F1 értékének függvénye. 370 Hz alaphfrekvencia fölött azonban mindenképpen két prominencia jelenik meg az észleletben, ezért az észleletet és a nyitottság érzetét nem a „teoretikus” F1 az  $f_0$ -hoz viszonyított távolsága, hanem az  $f_0$  és a  $2f_0$  távolsága befolyásolja elsődlegesen. Mégis, az F1 frekvenciaértéke még 370 Hz alaphfrekvencia fölött is befolyással van a zártságészleletre, hiszen az F1 függvényében az  $f_0$ -nál és  $2f_0$ -nál megjelenő prominenciák kétféle mintázatot eredmé-

nyezhetnek, ami – teoretikusan legalábbis – akár két eltérő észleleti hangzóminőséget is adhat. Az egyik mintázat az, melyben az F1 a 2f0-hoz (nem pedig az f0-hoz) van közel. Ez az, amit a nem első nyíltsági fokú magánhangzóknál tapasztalunk (beszélt ejtésben) 370 Hz-nél magasabb alapfrekvencián, és aminek – Traunmüller (1981) modelljének predikciója szerint – az eredménye a negyedik nyíltsági foknak megfelelő hangszínű /æ ɒ œ/ (és az egyszerűség kedvéért vegyük ide a Traunmüller kísérleteiben ötödik nyíltsági fokúnak minősített /a/ hangot is) megjelenése az észleletben, hiszen perceptiósan ez a mintázat az /æ ɒ œ a/ 370 Hz alatti megvalósulásaihoz hasonlít. A másik mintázat pedig az, melyben az F1 nem a 2f0-hoz, hanem az f0-hoz van közel, így annak a prominenciáját erősíti. Bár ezeknek a hangoknak az észleletére, mint említettük, Traunmüller (1981) modellje valójában nem ad predikciót, ám azt is láttuk, hogy a kísérletes eredményei szerint az ilyen mintázatot mutatott zárt ejtésű (első nyíltsági fokú) hangok megőrizték zárt hangszínüket. Mindehhez adalék, hogy mivel a 2f0 és a 3f0 (valamint a szekvenciálisan egyre magasabb vagy nagyobb sorszámú felharmonikusok) távolsága sosem nagyobb 3 Barknál, így a 2f0 felett megjelenő prominenciákat ismét csak a formánsok frekvenciái, nem pedig a felharmonikusok távolsága befolyásolja még 370 Hz alapfrekvencia fölött is.

A (magas alapfrekvencián) énekelt magánhangzók azonosításának vizsgálatában egyfelől az alapfrekvencia változásának, és így az F1 és f0 távolságának hatásaival, másfelől pedig az F1 : f0 hangolás befolyásával is számolnunk kell, mely akusztikai stratégiát akkor várjuk, ha az f0 frekvenciaértéke magasabb, mint az F1 beszédbeli megjelenésének értéke.

Az alapfrekvencia emelkedése az F1 : f0 hangolás megjelenése előtt és 370 Hz alapfrekvencia alatt azt eredményezi, hogy az F1 és f0 távolsága csökken, ami a tonotópiás modell predikciója szerint az egyes beszédhangok zártabbként azonosításához vezethet – ha az F1 és f0 különbsége átlépi a nyíltsági fokokat meghatározó határértékeket. Ha már itt, 370 Hz alapfrekvencia alatt fel lép az F1 : f0 hangolás (ami feltehetőleg csak az első nyíltsági fokba tartozó, alacsony F1 formánssal rendelkező /i: y: u:/ hangzóknál várható), az feltehetőleg nem okoz változást a zártság-észleletben, hiszen csak tovább csökkenti az F1 és f0 perceptiós távolságát (azaz csökkenti az észlelt F1 értékét), és zárt észleletet eredményez. A magasabb beszédbeni F1 formánssal rendelkező, a második, harmadik vagy negyedik nyíltsági fokba tartozó magánhangzók esetében 370 Hz alapfrekvencia alatt pedig nem várunk F1 : f0 hangolást.

370 Hz alapprofundencia fölött azonban elméletben két esetet különíthetünk el: azon magánhangzók esetét, amelyek első formánsát az énekesek a megemelt  $f_0$ -ra hangolják (melyek  $F_1$  értéke a beszédben alacsonyabb), valamint azokat, amelyekét nem. Azokban az esetekben, amikor 370 Hz fölötti alapprofundencián nincs  $F_1 : f_0$  hangolás, az észleletben elkülönülő két (az  $f_0$ -hoz és a  $2f_0$ -hoz eső) prominenciából a második az intenzívebb (hiszen ezt erősíti a magas frekvenciájú  $F_1$ ), így ezeket a hangokat az azonosítás során az elmélet szerint a negyedik nyíltsági fokú /a/-nak és /ɒ/-nak feleltetik meg a hallgatók. Azon hangok esetében azonban, melyek ejtése során az énekesek 370 Hz alapprofundencia fölött alkalmazzák az  $F_1 : f_0$  hangolást, az első, azaz az  $f_0$ -hoz eső prominencia lesz intenzívebb (nem pedig a  $2f_0$ -hoz eső). E hangok percepció s státusza azonban a tonotópiás modell alapján nem egyértelmű en bejósolható. Mivel elméletben ez a prominencia-mintázat nem jellemző a negyedik nyíltsági fokú hangok 370 Hz alatt ejtett megvalósulásaira, viszont hasonlít a 370 Hz fölött ejtett első nyíltsági fokú hangokra jellemző mintázatokra, feltételezhetjük, hogy a 370 Hz fölött ejtett,  $F_1 : f_0$  hangolt beszédhangokat inkább zártakként azonosítják az észlelők, de a magánhangzók minősége feltehetően nagyon bizonytalan. Mindemellert ha az  $f_0$  és  $2f_0$  prominenciája közti különbség megfelelően kicsi, az első nyíltsági fokú és negyedik nyíltsági fokú válaszok egyaránt valószínűek lehetnek az azonosítás során.

A jelen vizsgálat egyik célja, hogy korábban kipróbált és bemutatott technikák egy új kombinációjának segítségével, kikerülve az alul-mintavételezettség problémáját megállapítsa a magyar nyelv összes magánhangzóját érintő akusztikai változásokat, melyek az alapprofundencia emelését kísérik az éneklésben (különös tekintettel a magas alapprofundenciákra). Feltételezésünk szerint a magyar magánhangzókban is megjelenik a korábban, más nyelvek magánhangzóira leírt formánshangolási tendencia, azaz az első formáns az alapprofundenciára hangolása ( $F_1 < f_0$  helyzetben), de a hangolási tendencia megjelenésének „helye” (az alapprofundencia emelésének folyamatára vetítve) a magánhangzó zártságától, illetve a magánhangzó beszédben megjelenő  $F_1$ -értékétől függ.

A vizsgálat másik célja a magánhangzók azonosításának vizsgálata, valamint azok összefüggésbe hozása a magánhangzók produkcióját érintő változásokkal. A értekezés korábbi alfejezeteiben bemutatott vizsgálatok tanulságaira alapozva azt feltételezzük, hogy 900 Hz fölött a legtöbb hangzó megvalósulásra /a:/ válasz érkezik (az ejteni kívánt hangzó minőségétől függetlenül), ala-

csonyabb alaphfrekvenciákon azonban a magánhangzók az alaphfrekvencia emelésével egyre bizonytalanabb minősége miatt jóval változatosabb lesz a válaszok megoszlása. A hallási észlelés tonotópiás szerveződésén és az  $F1$  és  $f_0$  távolságának tonotópiás viszonyán alapuló elméletből kiindulva azt várjuk, hogy a modell szerint kritikus 370 Hz alatt (azaz a kísérletben vizsgált 349 Hz alaphfrekvencián, a  $f'$  zenei hangon és az alatt), az  $F1 : f_0$  hangolás megjelenése előtt a hangzók zártságának észleletét az  $F1$  és  $f_0$  észleleti távolsága határozza meg. Ugyanakkor azt is feltételezzük, hogy azon hangzók esetében, melyek ejtések az énekesek alkalmazzák az  $F1$  az  $f_0$ -ra hangolását, Traunmüller tonotópiás elméletének értelmében a hallgatók jó arányban azonosíthatnak zárt hangzókat is, valamint megjelenhet az egyes ejteni szándékozott középső nyelválású (második nyitásfokú) beszédhangok zártakként való felismerése is.

## 2.5.4. Anyag, módszer és kísérleti személyek

### 2.5.4.1. Felvételkedészítés és akusztikai elemzések

A kísérletben három profi szoprán énekes ( $S1$ ,  $S2$ ,  $S3$ , lásd a 7. táblázatot) énekprodukciját vizsgáltuk a kilenc hosszú, vagy hosszan ejthető magyar magánhangzó /ɒ a: ɛ e: i: o: ø: u: y:/ ejtése közben hat énekelt alaphfrekvencián (három oktáv hangterjedelemben) 175 és 988 Hz között (az  $f$ ,  $h$ ,  $f'$ ,  $h'$ ,  $f''$ ,  $h''$  zenei hangokon, lásd a 6. táblázatban) és beszéd módban.

7. táblázat: A kísérletben résztvevő énekesek

Énekes kódja	Életkor	Hol tanult?	Milyen formában, hol énekel?	Mióta énekel?
$S1$	28	Liszt Ferenc Zene-művészeti Egyetem	Szólistaként, kórusban, énektárnaként is dolgozik	14 éve
$S2$	45	Liszt Ferenc Zene-művészeti Egyetem	Szólistaként, kórusban, énektárnaként is dolgozik	Több, mint 30 éve
$S3$	28	Magánénektanárnál	Szólistaként, kórusban	14 éve

Ahogy eddig is, mivel a vizsgálatban a paraméterek szigorú kontrollálása, valamint a felvételi és elemzési protokoll sajátosságai miatt csak kitartott hangokat vizsgáltunk, öt fonológiai hosszúsági pár rövid tagja (az /i o ø u y/ magánhangzók) kimaradt az elemzésből. Ennek ellenére úgy tekinthetjük, hogy a jelen vizsgálati anyag reprezentálja a teljes magyar magánhangzókészletet,

hiszen az itt kimaradt hangokat hagyományosan csak időtartamukban tartjuk a pár hosszú tagjától eltérőnek a magyarban (l. pl. Gósy 2004). Az énekesek minden magánhangzót minden alapfrekvencián kétszer ejtettek ki, így összesen ( $3 \text{ énekes} \times 9 \text{ magánhangzó} \times 7 \text{ alapfrekvencia} \times 2 \text{ ismétlés} =$ ) 378 stimulust vizsgáltunk. Minden alapfrekvencián, minden kilencelemű magánhangzócsoporthoz felvétele előtt az énekesek referenciahangot hallottak fülhallgatón keresztül, így biztosítottuk a pontos intonációt.

A vizsgálat módszereiben az énekelt magánhangzók formánsainak meghatározásához két, korábban már sikerrel használt elemzési technikát ötvöztünk. Az első a magas alapfrekvenciájú zöngé helyettesítése és a toldalékcső „újramintavételezése” volt egy mély alapfrekvenciájú, nagy felhangsűrűségű külső gerjesztésű jel segítségével. A másik pedig a beszédjel (azaz az énekesek hangjának) inverz szűrése volt, melyhez EGG-t, illetve dEGG-t is használtunk.

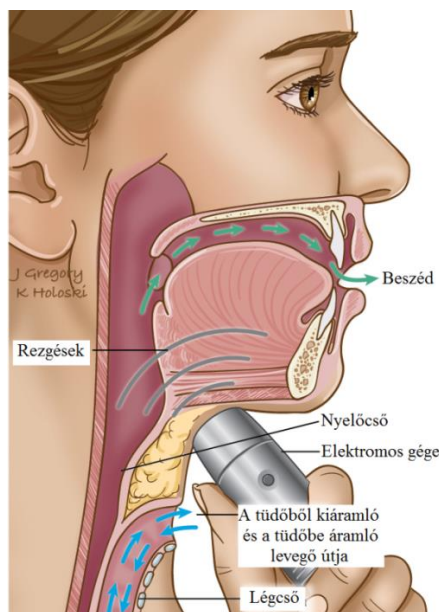
A mérési módszereknek megfelelően a felvételek során két jelet rögzítettünk párhuzamosan, két csatornán. Az első csatornán az énekes száját elhagyó hangot rögzítettük (audió) egy gömbkarakterisztikájú mikrofonnal, melyet 30 cm távolságra helyeztünk el az énekes szájától. (A mikrofont a szájnyílástól egy bukósisakra szerelt pálcá tartotta állandó távolságra, melyet az énekesek a felvétel során viseltek.) A másik csatornán pedig a hangszalagok mozgását rögzítettük egy elektrolottográf<sup>24</sup> segítségével.

Ahhoz, hogy a toldalékcsővön áthaladó, a szájat elhagyó kimeneti jelben is mérhető rezonanciákat kapjunk, az egyes hangok ejtésére jellemző artikulációs konfigurációt az énekes saját fonációja mellett egy elektromos gégeből (vagy más néven elektrolarynxból) származó külső gerjesztésű, mély alapfrekvenciájú, nagy felhangsűrűségű jellel is „mintavételeztük”. Az elektromos géget szokásosan a hagyományos módon valamilyen okból (általában részleges vagy egészesleges gégeeltávolítás miatt) zöngét képezni nem tudó személyek használják. Az eszköz mély, berregő hangot bocsát ki, melyet használója a gégeszint felett a nyak szöveteihez szorítva a toldalékcsőbe juttathat, a rezgés pedig (a zöngéhez hasonlóan) a garatüregen át a szájnyílás felé terjed tovább,

---

<sup>24</sup> A vizsgálatban egy Laryngograph® márkájú elektrolottográfot és az ahhoz tartozó *Speech Studio* szoftvert használtunk (Laryngograph, Laryngograph® Ltd, Wallington, Greater London, UK). Mind az elektrolottográfot, mind pedig a felvételek helyszínéül szolgáló csendesített szobát az MTA Nyelvtudományi Intézetének Fonetikai osztályán működő Kempelen Farkas Beszédkutató Laboratórium bocsátotta rendelkezésünkre, amit ezúton is hálásan köszönünk a felvételt készítésben segítséget nyújtó Grácz Tekla Etelka és Beke András segítségével egyetemben.

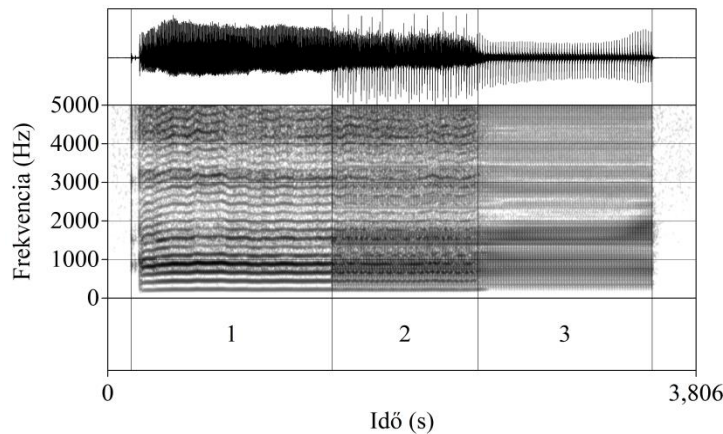
mígnem végül távozik a szájnyíláson keresztül. A száját elhagyó, az elektromos gége gerjesztette hang spektrumát a toldalékcső üregei – a beszédhangképzéssel analóg módon – rezonátortevékenyséjük révén módosítják, így a kimenetben beszédhangszerű hangok észlelhetők és rögzíthetők (l. 29. ábra).



29. ábra: Az elektromos gége működési elve gégeltávolításon átesett páciensről készült keresztmetszeti ábrán szemléltetve. Az elektromos gége keltette berregő hang a nyak szövetein keresztül a toldalékcsőbe jut, majd a garatüregen, és végül pedig a szájüregen is áthaladva távozik a szájnyíláson keresztül. A hang spektrumát a toldalékcső üregei rezonátortevékenyséjük révén módosítják, így a kimenetben beszédhangszerű hangok észlelhetők és rögzíthetők (<http://www.headandneckcancerguide.org/> alapján).

Az így keletkező akusztikai jel elemzése tehát alkalmas arra, hogy benne az artikulációs konfigurációra az ejtés közben jellemző formánsfrekvenciákat meghatározzuk. Az elektromos gégével történt felvételekhez az énekeseket kitartott hangok éneklésére kértük. A felvétel folyamatait a 30. ábrán szemléltetjük, és ez alapján az ábra alapján mutatjuk be. Az ábra az egyik beszédben ejtett /a:/ hang spektrogramját mutatja a kétféle mintavételezési eljárással készült felvételben.

Az énekesek feladata az volt, hogy a magánhangzó ejtése közben pár másodpercig énekeljenek természetesen, azaz képezzenek zöngét (30. ábra, 1-es szakasz), majd kapcsolják be az ekkor már a nyakukhoz szorított elektromos géget (30. ábra 2-es szakasz), végül pedig merevítsék meg az artikulációjukat, és hirtelen hagyják abba a zöngéképzést úgy, hogy a hangrést a levegő visszatar-  
tásával bezárják (30. ábra 3-as szakasz).



30. ábra: Az /a:/ hangzó beszédbeli ejtéséről készült keskenysávú spektrogram S1 énekes ejtésében. A spektrogram a felvételek során használt kétféle „mintavételezési” eljárást (valamint azok átlapolását) szemlélteti: (1): az énekes zöngét képez; (2) az énekes zöngéképzés közben bekapcsolja a gégeszint felett a nyakához illesztett elektromos géget; (3) az énekes hirtelen abbahagyja a zöngéképzést, visszatartja a levegőt, bezárja a hangrést, így a felvétel már csak az elektromos gégeből származó, az énekes száján keresztül távozó hanghullámokat rögzíti.

Ilyen módon a (2)-es számmal jelölt szakaszon az énekes toldalékcsovét mind saját zöngéje, mind pedig az elektromos gége hangja is mintavételezte, míg a (3)-as számmal jelölt szakaszon már csak az elektromos gége a toldalékcső által módosított hanghullámai hagyták el a száj-  
üreget. Mivel a hangrést az énekesek ez utóbbi, azaz a (3)-as szakaszban bezárták, így a szubglottális területek rezonátortevékenysége itt nem lehetett hatással a kimenő jel spektrumára. Az ilyen módon ejtett magánhangzókat rögzítettük a bukósisakra rögzített mikrofon segítségével,



majd elemeztük a Praat szoftverben (Boersma–Weenink 2009). Az elektromos gége használatát a felvételek elkészítése előtt a szükséges ideig minden énekessel gyakoroltuk.

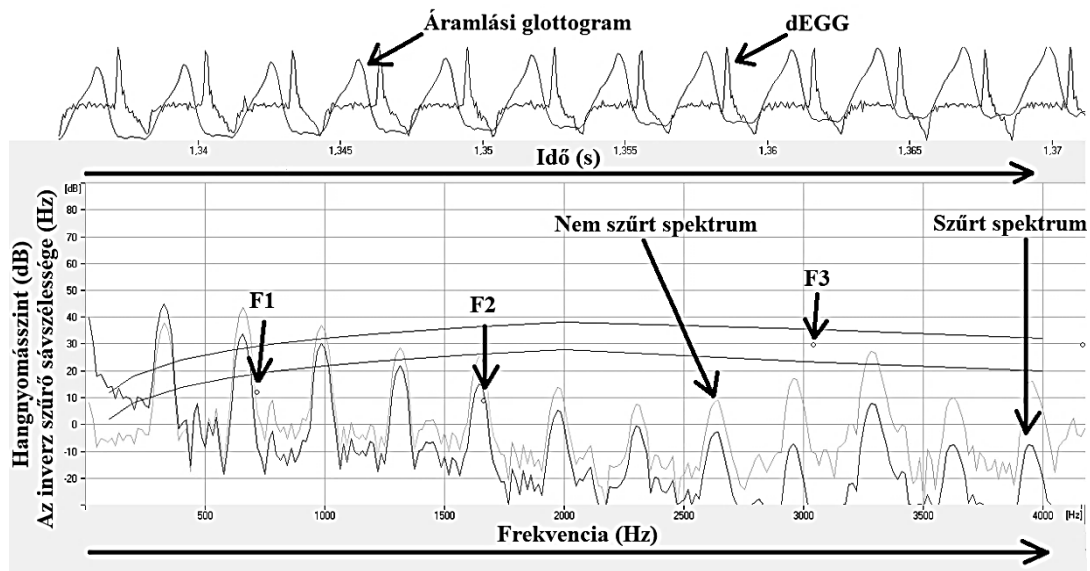
A kísérletben a természetes módon produkált (az ének és beszéd módban) rögzített magánhangzókat inverz szűrés segítségével elemeztük a DeCap szoftverben (Granqvist, é. n.)<sup>25</sup>. A szűréshez a beszédjelet és az EGG-jelet tartalmazó hangsávokat először időben illesztettük. Mivel a két jelfelvételi pont, azaz az EGG-elektrodák (vagyis a hangszalagok) és a mikrofon közti távolság nem nulla, ezért a két sávot egymáshoz képest időbeli elcsúszásban rögzítettük. Ezt a (közel 45 cm-es) távolságot a jelek szinkronizálásához úgy kompenzáltuk, hogy kiszámítottuk, hogy a hang terjedési sebességéből, valamint a két pont közti távolságból következően mekkora idői eltolódásban rögzítettük a két csatornát, és a kapott értékkel eltoltuk (azaz „előrébb hoztuk”) az EGG csatornát. (Mivel a toldalékcső felnőtt nőknél átlagosan 15 cm, és ehhez adódott hozzá a mikrofon távolsága az ajkaktól, így 45 cm távolsággal számoltunk.) Az eltolás mértékét matematikai úton ( $t = \Delta s / v = 45 \text{ cm} / 35\,000 \text{ cm/s} =$ ) 1,3 ms-ban állapítottuk meg, amit a beszéden végzett jelek inverz szűrése után 2 ms-ra módosítottunk (úgy, hogy a dEGG-ben detektálható záródási csúcs, valamint az áramlási glottogram zárszakaszának kezdete tökéletes szinkronba kerüljön).

A DeCap szoftver manuális inverz szűrést tesz lehetővé. A program egyszerre kezeli az EGG-t és a mikrofon által felvett audiojelet: az EGG-t (és deriváltját, a dEGG-t) a felület felső ablakában hullámformaként jeleníti meg, a mikrofon által rögzített jelet pedig egyszerre hullámformaként (szintén a felső ablakban), és az alsó ablakban spektrumként is látjuk (31. ábra). Az elemzéshez a program először integrálja a hangnyomás-idő függvényt, azaz levegőáramlás-idő függvénné alakítja (abból kiindulva, hogy a levegőáramlás-idő függvény deriváltja a hangnyomás-idő függvény). Ezt a jelet szűrjük az elemzés során. Emellett pedig az EGG is (tetszés szerint) deriválható, tehát a dEGG jel is megjeleníthető. Az elemzésben a 31. ábrán F1, F2, F3 jelöléssel jelzett formánsokat, valamint (szükség szerint) a további, magasabb frekvenciájú és sorszámú formánsokat manuálisan állítjuk: vízszintesen állítjuk a formáns frekvenciáját, függőlegesen pedig a formáns sáv szélességét. A formánsok (azaz a szűrők) manuális hangolásának megfelelően pedig mind a hullámformát, mind pedig a spektrumot valós időben módosítja a program (a klasz-

---

<sup>25</sup> Az elemzéseket a KTH Royal Institute of Technology, Speech Music and Hearing tanszékén végeztem prof. emer. Johan Sundberg, dr. Sten Ternström és dr. Svante Graqvist segítségével, amit ezúton is hálásan köszönök.

szikus egyenleteknek megfelelően, l. Fant 1960), tehát a szűrés eredménye azonnal látható lesz (miközben pedig a nem szűrt, eredeti jelet is megjeleníthetjük, az ábrán ez látható halványabb színnel az alsó, FFT-ablakban).



31. ábra: Az inverz szűrés menete a DeCap programban. A felső ablakban az EGG, a dEGG (sötétebb színnel) valamint a szűrt áramlási glottogram látható (a hangnyomás-idő függvény integráltja, világosabb színnel). Az alsó ablakban az audiofelvétel eredeti (világos vonal) és szűrt (sötétebb vonal) spektruma látható, valamint a beállított szűrők (apró körök; F1, F2, F3) frekvenciája és sáv szélessége (Sundberg et al. 2013 alapján).

Az inverz szűrés, azaz a szűrők manuális hangolásra során a cél 1. a dEGG-ben azonosítható záródási csúcs és az áramlási glottogramon beazonosítható zárszakasz kezdetének<sup>26</sup> szinkronizálása, 2. a zárszakasz-hullámforma fodrozódásainak megszüntetése, és 3. a spektrum meredekségének kiegyenlítése a lokális maximumok és minimumok lehetőség szerinti teljes kisimításával (vö. Hertegård–Gauffin 1993, Sundberg 2011, 2013). A szűrők helyes beállításával az áramlási glottogram a glottiszon áthaladó levegő (elméleti) hullámformáját mutatja (vö. 28. ábra). A 31. ábrán látható áramlási glottogram hullámformáján a szűrés utáni állapotot láthatjuk (egy jól beál-

<sup>26</sup> Sundberg (2011, 2013) erre a pontra az MFDR megnevezéssel (*Maximum Flow Declination Rate*) hivatkozik, mely kifejezés az áramlási glottogram maximális meredekségére utal a záródási (azaz deklinációs) szakaszban.

lított szűrősor alkalmazása esetén). Az alsó, FFT-ablakban látható közel párhuzamos és megközelítőleg vízszintes vonalak (a „kerítés”) az elemzést segítő az adott frekvencián reális formáns-sávszélességeket jelölik. Ezekről, női beszélő lévén gyakran eltértek a kapott adatok, a legtöbb esetben ugyanis a javasoltnál nagyobb sávszélességeket mértünk az első két formáns esetében (azaz a formások értékét mutató jelölők a „kerítés” alatt helyezkedtek el, amint azt a 31. ábrán is látjuk).

Az inverz szűréssel kapott formánsértékeket összevetettük az elektromos gégével felvett (a 30. ábrán a (3)-as számmal jelölt) hangzórészletek elemzésén keresztül nyert adatokkal. A kétféle módszerrel kapott adatok eltérése esetén ellenőriztük a lehetséges okokat, és javítottuk a hibákat, egyezés esetén pedig elfogadtuk a mérést.

#### 2.5.4.2. Percepció vizsgálat

A vizsgálat második felében a felvett hangmintákon percepció tesztet végeztünk 21 magyar anyanyelvű, ép halló, egészséges kísérleti személy részvételével (12 nő, 9 férfi, 26 és 45 év közöttiek, életkoruk átlagosan 29 év) egy, a Praat programban (Boersma–Weenink 2009) írt és futtatott szkript segítségével. A három énekes által 6 alaphangfrekvencián (175 és 988 Hz között, l. 6. táblázat) felénekelte és beszéd módban ejtette /p a: e i: o: ø: u: y:/ hangzókat a percepció teszthez némileg manipulálnunk kellett: a Wavesurfer program (Sjölander–Beskow 2000) segítségével kiegyenlítettük az egyes felvételek esetlegesen eltérő hangerősségét, valamint elimináltuk a felvételekből az elektromos zöngé bekapcsolását követő szakaszt (lineáris fokozatos kiűzés effektus segítségével) úgy, hogy az egyes stimulusok megközelítőleg egyenlő (1,8–2 s) hosszúságúak legyenek. A tesztben az így módon egységesített hanganyagot minden adatközlőnek kétszer játszottuk le adatközlőnként eltérő, randomizált sorrendben (így az esetleges fáradás következtében megjelenő hatások eloszlottak a stimulusok között, valamint azt is kontrolláltuk, hogy ne lépessen fel az egymást követő stimulusok szisztematikus egymásra hatása az észleletben).

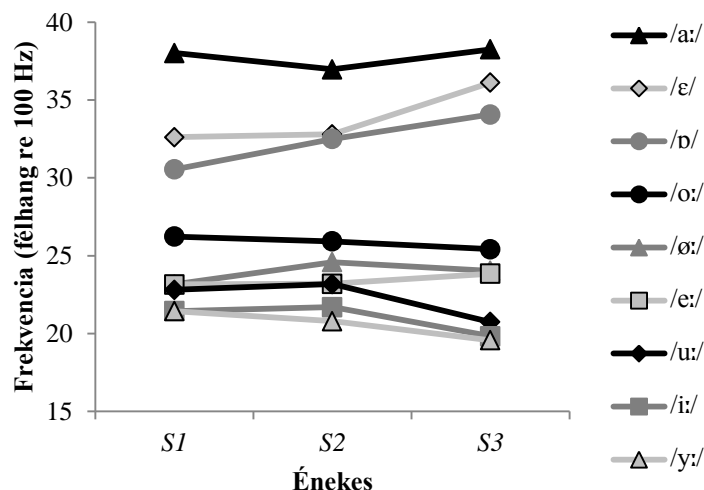
A résztvevők feladata mindkét tesztfázisban ugyanaz volt: ki kellett választaniuk a képernyőn ortografikus alakban megjelenő kilenc magyar magánhangzó közül azt, amelyiket hallani vélték. A válaszlehetőségek a hosszú (vagy fonémakategória-váltás nélkül hosszan ejthető) magyar ma-

gánhangzók voltak: /ɒ a: ɛ e: i: o: ø: u: y:/. A teszteket csendes körülmények között vettük fel, a hanganyagokat zárt fejhallgatón keresztül hallották a résztvevők.

### 2.5.5. Eredmények

Mivel az *S2* kódjelű énekes a felvételek során (fáradtság miatt) nem tudta megfelelő hangszínezettel végigénekelni a stimulusokat a *h''* (988 Hz) zenei alapfrekvencián, ezért az eredményeket szemléltető ábrákon az *S2* kódjelű adatsorokról a megfelelő adatok hiányoznak. Ugyanezen okból a 34. ábrán látható magánhangzótér-változásokat is a legmagasabbnál eggyel alacsonyabb (*f''*, azaz 698 Hz) alapfrekvencián szemléltetjük, ahol mind a három énekes összes adata rendelkezésre állt.

A 32. ábra a beszélt módban ejtett magánhangzó-megvalósulások első formánsának értékét mutatja (félhangokban). Látható, hogy a három énekesnek a beszéd módban ejtett magánhangzói hasonló F1-értékkel realizálódtak. A beszélők (énekesek) közti kisebb-nagyobb eltérések ellenére a nyíltabb hangzók között nagyjából egyértelműen felállítható sorrendet látunk: a legnyíltabban, legmagasabb F1-értékkel az /a:/ hangot ejtették az énekesek, melyet rendre az /ɛ/ és /ɒ/ követett (a nyíltságot természetesen itt szigorúan akusztikailag értve). A zártabban ejtett hangzók közti különbségek már nem ilyen egyértelműek, de az minden énekes esetében megfigyelhető, hogy az alsó nyelvéllátásúaknál valamivel alacsonyabb F1-értékkel a középső nyelvéllátású magánhangzók realizálódnak (az /o:/, az /ø:/ és /e:/), míg a legalacsonyabb F1-értékkel a felső nyelvéllátású magánhangzók (az /u:/, az /i:/ és az /y:/).



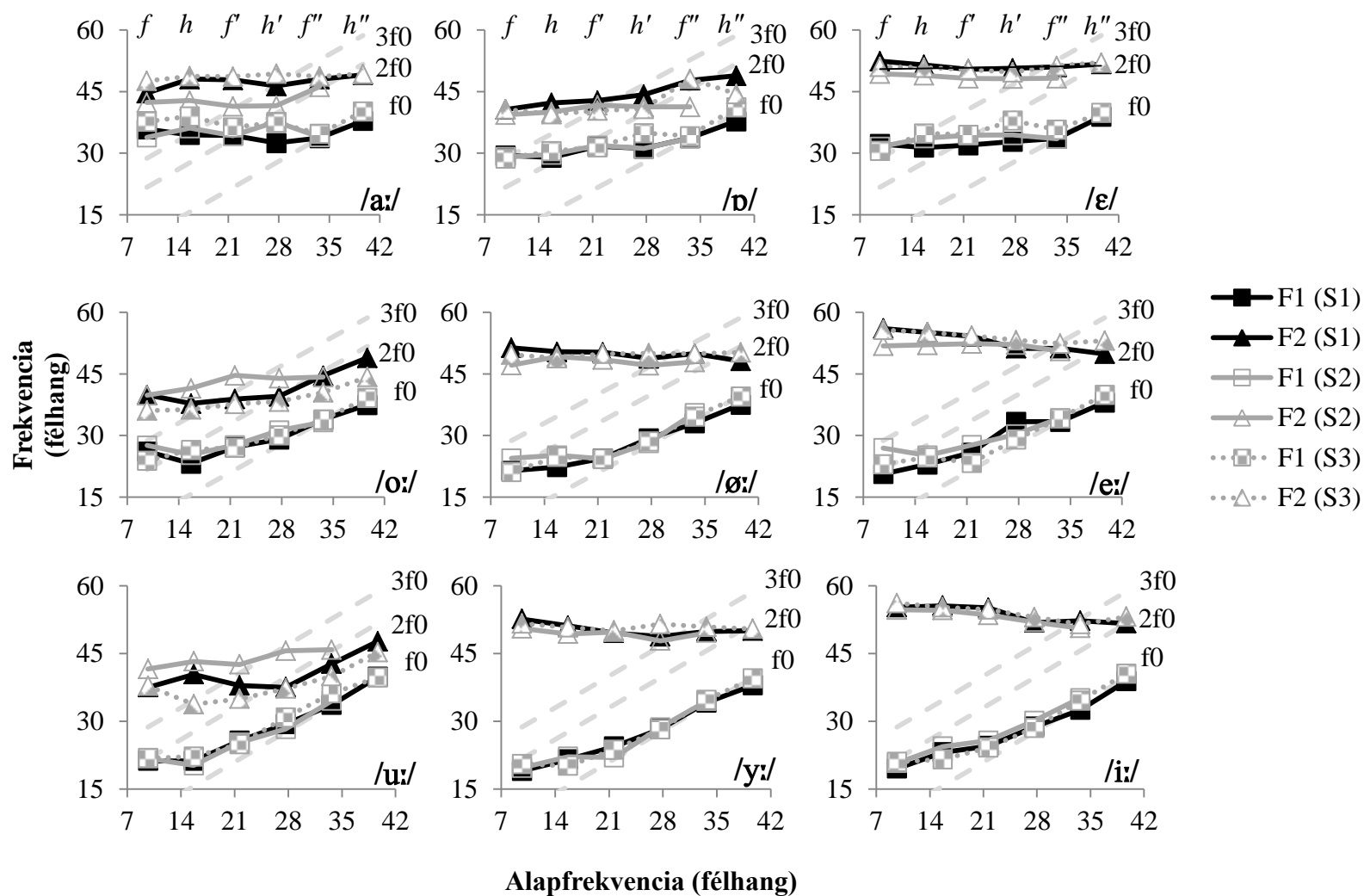
32. ábra: A beszédben ejtett magánhangzók első formánsa a három énekes ejtésében

#### 2.5.5.1. Az énekelt magánhangzók első formánsának alakulása

A 33. ábra a kilenc vizsgált magánhangzó első és második formánsának értékeit mutatja ( $y$  tengely) az alapfrekvencia függvényében ( $x$  tengely), félhangokban mérve (a lineáris Hertz értékeket az inverz szűrés elvégzése után konvertáltuk félhangokká, minden esetben 100 Hz frekvenciához viszonyítva) a három énekes ejtésében külön-külön. Az ábrákon világos színű szaggatott vonallal jelöltük az alapfrekvenciát és annak egész számú többszöröseit, azaz a félhangokat vagy felharmonikusokat ( $f_0$ ,  $2f_0$ ,  $3f_0$  stb.). Az énekesek közti kisebb mértékű variabilitás ellenére egyértelmű tendenciákat találtunk: mindhárom énekes ejtésében megjelent az első formáns az alapfrekvenciára (vagy ahhoz közel eső frekvenciaértékre) hangolása akkor, ha az  $f_0$  megközelítette az  $F1$ -nek a beszédben az adott magánhangzóra jellemző értékét (vö. 32. ábra). Attól a ponttól kezdve, ahol az  $f_0$  az (alapfrekvencia-emelés során) elérte az  $F1$ -nek a beszédben jellemző régióját, az  $f_0 : F1$  hangolást minden énekes alkalmazta minden magasabb  $f_0$  és minden magánhangzó esetében. A hangolási tendencia megjelenésének alsó határa tehát az  $F1$ -től, azaz az adott hangzó nyelvállásfokától/nyitottságától függött: a legmagasabb ( $h''$ , 988 Hz) a legnyitottabb /a: ɒ ε/ esetében volt, ennél alacsonyabban tapasztaltuk ( $h'$ , 494 Hz) a zártabb vagy magasabb nyelvállásfokú /o: ø: e:/ esetében, míg a legalacsonyabban ( $f'$ , 349 Hz) a legzártabb, felső nyelvállású /u: y: i:/ esetében láttuk megjelenni. A hangolási tendencia megjelenése, azaz a fenti alapfrekvenciák

elérése előtt az F1 frekvenciaértéke az  $f_0$ -tól függetlenebbnek mutatkozott, és a beszédhez közeli frekvenciaregióban realizálódott.

Az F1 :  $f_0$  hangolási tendencia felső határa tekintetében az eredményeink nem egyeznek Joliveau és munkatársainak (2004) eredményeivel, akik a hangolási tendencia hiányát találták a kísérletükben vizsgált ajakkerekítéses hangok (/ɔ u/) esetében. A jelen vizsgálatban ugyanis az F1 :  $f_0$  hangolásnak nem volt felső határa, sem az ajakréses, sem az ajakkerekítéses hangoknál. Pontosabban fogalmazva a zenei  $f''$  hang (988 Hz alapfrekvencia) elérésével a résztvevő három énekes nem érte el az F1 :  $f_0$  hangolási tartományuk felső határát, ugyanis nem láttuk a hangolás sikertelenségét (míg Joliveau és munkatársainak kutatásában a vizsgált alapfrekvencia felső határa valahol 1 kHz fölött, feltehetőleg 1046 Hz-en, azaz a „magas C-n” volt, ahol az ajakkerekítéses hangzók első formánsa már az  $f_0$  értéke alatt volt mérhető).



33. ábra: A három szoprán énekes által ejtett beszédhangok F1 és F2 formánsfrekvenciái (y tengely) az alapfrekvencia (x tengely) függvényében. Minden jelölő egy énekes egy adott magánhangzójának egy adott alapfrekvencián ejtett megvalósulását mutatja. Világosszürke szaggatott vonalakkal az alapfrekvenciát és annak első két felharmonikusát jelöltük (rendre  $f_0$ ,  $2f_0$ ,  $3f_0$ ).

#### 2.5.5.2. Az énekelt magánhangzók második formánsának alakulása

Az F2 értéke az F1-nél jóval függetlenebbnek mutatkozott az alapfrekvenciától, különösen az elől képzett, valamint centrális képzésű /ε a:/ hangzók esetében. Az elől képzett /e: i: y:/ F2 értékében az F1 : f0 hangolásának megjelenésével kezdődően az f0 (és F1) emelésével párhuzamos, fokozatos, de enyhe csökkenést tapasztaltunk az alapfrekvencia emelésével. A hátul képzett /ɒ o: u:/ F2-értéke esetében azonban az előbbieknél jelentősebb mértékű összefüggést láttunk az alapfrekvenciával, ugyanis e hangok második formánsa (az F1 : f0 hangolás megjelenésével azonos alapfrekvencián kezdődő és az f0 emelésével párhuzamos) fokozatos és nagymértékű növekedést mutatott.

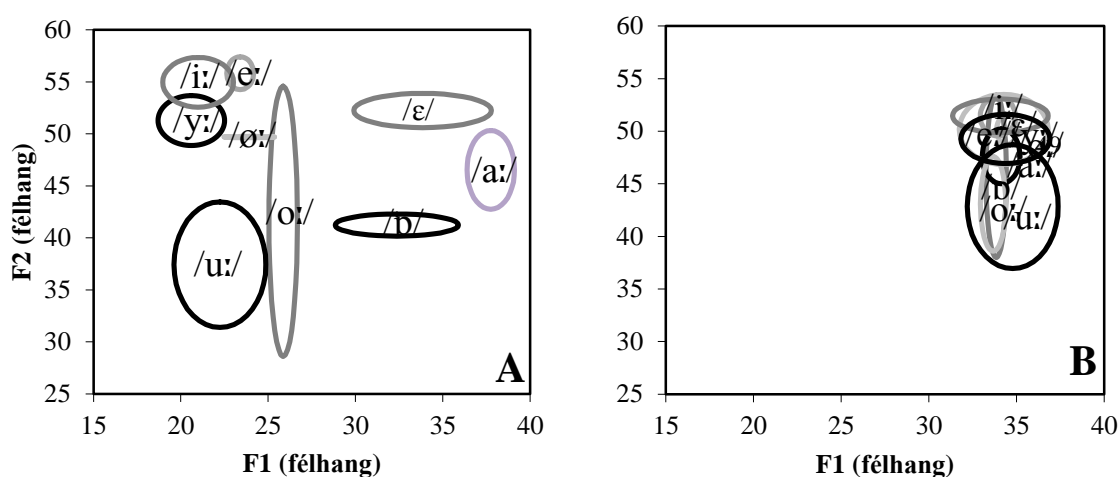
Az F2-értéknek változására jellemző tendenciák az akusztikai magánhangzótér kihasználtsága szempontjából azt jelentik, hogy az elől és hátul képzett magánhangzók az alapfrekvencia emelésével az F1 : f0 hangolás megjelenésétől (azaz nagyjából az  $f'$  és  $h'$  zenei hangoktól, azaz körülbelül 300-500 Hz alapfrekvenciától) kezdve egyre kevésbé különültek el egymástól. Ezt a konvergálást azonban elsősorban a veláris hangok fokozatosan változó, egyre palatálisabb képzése okozta. Az F2 változása tehát a vízszintes nyelvhelyzet függvényében eltérő módon volt jellemző a magánhangzókra, de az F2 változásában (azaz az elől-hátulképzettség változásában) a nyíltsági foknak is szerepe volt: a palatálisok között az első és második nyíltsági fokú /e: i:/, míg a velárisok között az ugyanezen nyíltsági fokokba tartozó /o: u:/ F2-értéke függött a leginkább az f0-tól, és változott a legjelentősebben az alapfrekvencia emelésével.

#### 2.5.5.3. Az akusztikai magánhangzótér az alapfrekvencia függvényében

Az akusztikai magánhangzótér az előző alfejezetekben már említett magas alapfrekvencián tapasztalt változásait a 34. ábrán szemléltetjük. Az ábrán szereplő koordináta-rendszerek két tengelyét az F1 és F2 magánhangzó-formánsok adják (ezeket rendre az  $x$  és az  $y$  tengelyen ábrázoltuk), így tehát az  $x$  tengely a nyíltsági fokkal, a függőleges nyelvhelyzettel, az  $y$  tengely pedig a veláris és palatális képzéshellyel, azaz a vízszintes nyelvhelyzettel függ össze. Az  $x$  tengelyen balról jobbra haladva találjuk rendre az első, második, harmadik és negyedik nyíltsági fokú magánhangzókat, míg az  $y$  tengelyen fentről lefelé haladva a veláris és palatális magánhangzók külö-



nülnek el egymástól. Az *A* panel a beszéd módban ejtett magánhangzók eloszlását, a *B* panel pedig a 698 Hz alapfrekvencián, azaz a zenei  $f''$  hangon énekelt magánhangzók megvalósulásait mutatja az akusztikai magánhangzótérben a három énekes ejtésében: az ábrákon az egyes magánhangzó-megvalósulásokat reprezentáló ellipszisek középpontjainak koordinátáit az előzőekben ismertetett F1- és F2-értékek átlagával, míg az ellipszisek sugarát a megfelelő adatok kétszeres szórásával határoztuk meg.



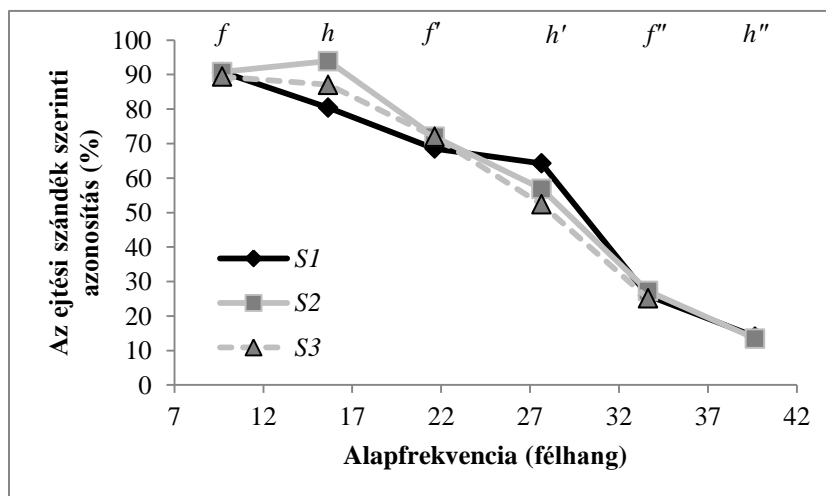
34. ábra: A három énekes által énekelt magánhangzók megvalósulásának eloszlása a (pszicho)akusztikai magánhangzótérben a beszédben (*A* panel) és az  $f''$  zenei alaphangon, 698 Hz alapfrekvencián (*B* panel). A magánhangzók megvalósulásait reprezentáló ellipszisek középpontját az F1 és F2 formánsok értékeinek átlagával, az ellipszisek sugarát pedig a megfelelő adatok kétszeres szórásával határoztuk meg.

Az *A* panelen látható, hogy a beszédben megvalósult magánhangzók az érzetileg transzformált (félhangokat, azaz logaritmikus skálát alkalmazó) pszichoakusztikai térben jól elkülönülnek egymástól, az ellipszisek közti fedés, tehát a beszédhangok megvalósulásai közti átfedés csak kismértékben jellemző. A *B* panelen ezzel szemben a magánhangzótér zsugorodását és a nyílt és centrális ejtés felé, azaz az /a:/ felé történt elmozdulását figyelhetjük meg az  $f''$  zenei hangon (ezt az ábrán az ellipszisek horizontális irányban jobbra, valamint vertikális irányban középtájra történt elmozdulása mutatja).

Bár az  $f''$  zenei hang nem számít kifejezetten magasnak a szopránok hangtartományát tekintve, már ezen az alaphangfrekvencián is tapasztalható volt a veláris és palatális hangok elkülönítésének drasztikus csökkenése, és a hangzók akusztikai nyitódása, amit az alaphangfrekvencia extrém emelésével vártunk. Ugyanakkor az is látható, hogy az elől és hátul képzett hangok ezen a közepesen magas alaphangfrekvencián még valamivel nagyobb mértékben elkülönülnek egymástól (mint például a különböző nyíltsági fokú hangzók).

#### 2.5.5.4. A percepció adatok és a percepció adatok összefüggései az akusztikai eredményekkel

Annak megfelelően, hogy a magánhangzók artikulációs és akusztikai célja az alaphangfrekvencia emelésével változik, azt várjuk, hogy a magánhangzók azonosítása is egyre kevésbé egyezik meg az ejtési szándék szerinti magánhangzó-minőséggel. A három énekes által énekelt kilencféle magánhangzóra kapott, az ejtési szándéknak megfelelő átlagos azonosítási arányokat a 35. ábrán mutatjuk be.



35. ábra: A három énekes ( $S1$ ,  $S2$ ,  $S3$ ) által énekelt magánhangzók azonosítása az alaphangfrekvencia függvényében

Az ábrán az egyes jelölők egy-egy énekes egy adott alaphangfrekvencián ejtett összes magánhangzójára kapott ( $21 \text{ adatközlő} \times 9 \text{ magánhangzó} \times 2 \text{ ismétlés} =$ ) 378 darab válaszból számított szá-

zalékos arányt mutatják. Így például azt látjuk, hogy a legalacsonyabb énekelt alapfrekvencián (a zenei  $f$  hangon, azaz 175 Hz-en) a mindhárom énekes énekelt magánhangzóra adott válaszok körülbelül 90-90%-a volt az ejtési szándéknak megfelelő, azaz az ejtési szándékkal megegyező hangminőséget megjelölő válasz. Az adatokból kiderül, hogy az énekelt magánhangzókra általánosan mindhárom énekes ejtésében ugyanolyan észlelési tendenciák jellemzőek az alapfrekvencia függvényében: a magánhangzók az ejtési szándéknak megfelelő azonosítása az alapfrekvencia emelésével fokozatosan csökken, mígnem a legmagasabb  $h''$  hang, azaz 988 Hz alapfrekvencia elérésével megközelíti a véletlen (11%) szintjét. (Emlékeztetőül: mivel a legmagasabb alapfrekvencián  $S2$  énekes nem tudta végigénekelni az összes kísérleti stimulust, így ezen az alapfrekvencián csak  $S1$  és  $S3$  énekelt magánhangzóit teszteltük).

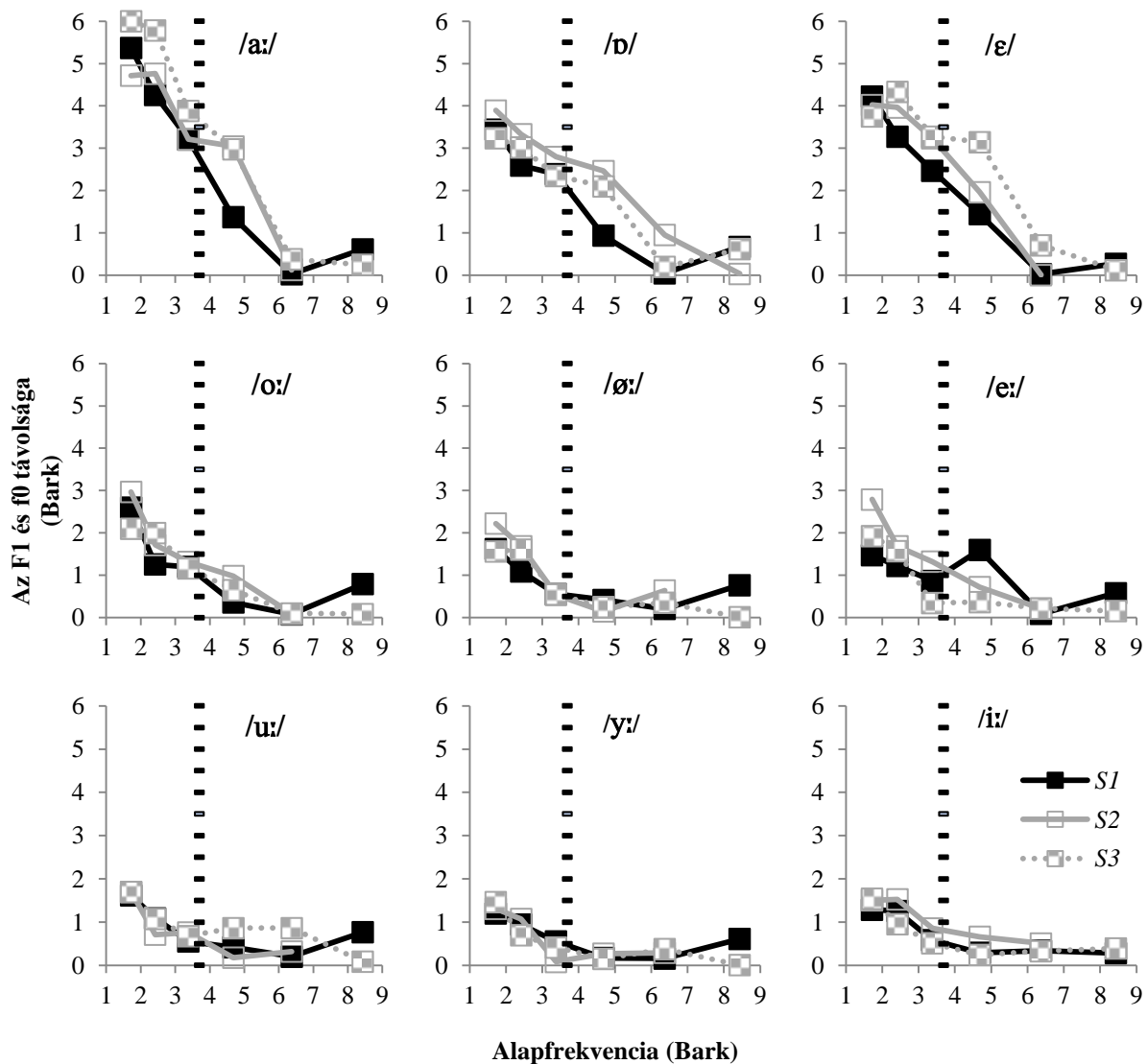
Az egyes magánhangzókra jellemző azonosítási tendenciák vizsgálatához érzetileg (Bark skálára) transzformáltuk az előzőekben az egyes hangokra mért  $F1$  és  $f0$  értékeket. Az  $F1$  és  $f0$  különbségét, mely Traunmüller (1981) szerint a zártság észleletét befolyásolja (legalábbis bizonyos alapfrekvencia alatt) a 36. ábrán szemléltetjük ( $y$  tengely) az alapfrekvencia függvényében ( $x$  tengely).

Traunmüller (1981) a rezonanciák és az alapfrekvencia tonotopikus viszonyára alapozó modelljének megfelelően a különböző nyíltsági fokú magánhangzók elkülönítése szempontjából a 370 Hz alapfrekvencia számít kritikusnak. Traunmüller modellje ugyanis elsősorban ez alatt a 370 Hz-es alapfrekvencia feltételez különböző nyíltsági fokú (azaz különböző nyíltsági fokúként azonosítható) magánhangzókat. Ennek megfelelően, a jelen vizsgálatban az itt vizsgált négyféle nyíltsági fok elkülönítésében a  $f'$  (349 Hz) és a  $h'$  (494 Hz) zenei alaphangoknak szentelünk kiemelt figyelmet (és elsősorban arra teszünk megfigyeléseket, hogy mi történik ez alatt és e fölött a frekvenciasáv fölött). Emlékeztetőül, a magánhangzók érzeti nyíltságát az  $F1$  és  $f0$  távolságából levezető modell szempontjából azért tekinthető fontos határnak a 349 Hz és 494 Hz alapfrekvencia, mert 349 Hz-ről 494 Hz-re lépve az  $f0$  és  $2f0$  közti távolság 2,99 Barkról 3,75 Barkra nő, tehát átlépi a 3 Bark perceptuális integrációs határt, mely alapvetően megváltoztatja az ezeken a frekvenciákon ejtett beszédhangok azonosíthatóságát.

Traunmüller elméletéből következően 349 Hz ( $f'$ ) alatt az  $/\text{ɒ}/$  és  $/\text{a:}/$  kivételével minden ejteni szándékozott hangzó első formánsának és alapfrekvenciájának frekvenciasávján csak egy promi-

nencia jelenik meg az észleletben – az /ɒ/ és /a:/ esetében azonban kettő, mert itt az F1 és f0 távolsága nagyobb, mint 3 Bark. A legzártabb /i: u: y:/ hangzók esetében tehát ezen az alapfrekvencia-tartományon azt vártuk, hogy ha az F1 értéke változatlan, és az F1 és f0 tonotopikus távolsága nem nagyobb, mint 1,2–1,5 Bark, a hangzókat zártakként, tehát az ejtési szándék szerint ismerik fel a tesztben részt vevők. A 36. ábra tanúsága szerint bár az /u:/ esetében a legalacsonyabb énekelte alapfrekvencián a megadottnál valamelyest nagyobb az F1–f0 távolsága, az /u: y: i:/ első formánsának és az f0-nak a távolsága a megadott tartományon belül marad a kritikus, 370 Hz-es alapfrekvenciáig (melyet az ábrán függőleges szaggatott vonal jelez). A 37. ábrán látható azonosítási adatok szerint az észlelésben a modellből bejósolható tendenciákat tapasztaltuk. Az ábrán ugyanis azt látjuk, hogy míg az /i:/-t és /y:/-t a kritikus 370 Hz alapfrekvencia alatt, azaz az első három vizsgált alapfrekvencián a legnagyobb arányban első zártsági fokú hangzóként (tehát /i:/-ként és /y:/-ként) azonosították a tesztelők, az /u:/-t bár főként az ejteni kívánt hangzó nyíltsági fokának megfelelően észlelték (tehát /u:/-ként), mégis kisebb számban, nyitottabb hangzóként (/o:/-ként) is azonosították.

Az akusztikai adatok szerint a kritikus 370 Hz alapfrekvencia alatt a zárt /u: y: i:/ magánhangzók esetében 349 Hz-en (*f'*) már fellép az F1 : f0 hangolás. Traunmüller modellje alapján azt feltételezhetjük, hogy a hangolás következtében megváltozott akusztikum észlelete ugyanolyan marad, mintha nem lépne fel hangolás (legalábbis 370 Hz alatt), mert az F1 : f0 hangolás következtében az első észleleti prominencia frekvenciája csökken, és ez elviekben csak erősíthet a hangzók első zártsági fokú észleletén. A percepció adatok szerint ez a modellből levezethető predikció helytálló. Az azonosítási adatokban ugyanis azt látjuk, hogy a hangolás nem hoz drasztikus változást az azonosításban: mind a palatális, mind a veláris hangzók esetében az eggyel alacsonyabb alapfrekvencia szerinti arányokat látjuk, tehát az /y: i:/ első zártsági fokú hangokként való azonosítását, valamint az /u:/-ra az /u:/ mellett megjelenő kisebb számú /o:/ válaszokat.



36. ábra: Az F1 és f0 érzeti távolsága (y tengely) az alaphangmagasság (x tengely) függvényében a három énekes által ejtett kilenc magánhangzóban. A függőleges szaggatott vonalak a Trnumüller (1981) modellje által jósolt kritikus, 370 Hz alaphangfrekvenciájú alaphangmagasságot jelölik.

A nagyobb nyíltsági fokú magánhangzók /o: ø: e:/ esetében 370 Hz alatt még nem lép fel F1 : f0 hangolás, így ezeknél a hangzóknál csak az alaphangfrekvencia emelése miatt csökkenő F1 és f0 közti tonotopikus távolsággal kell számolnunk mint a zártságészleletet befolyásoló paraméterrel. A modell értelmében e hangok esetében 370 Hz alatt az f0 és F1 frekvenciasávján egy észle-

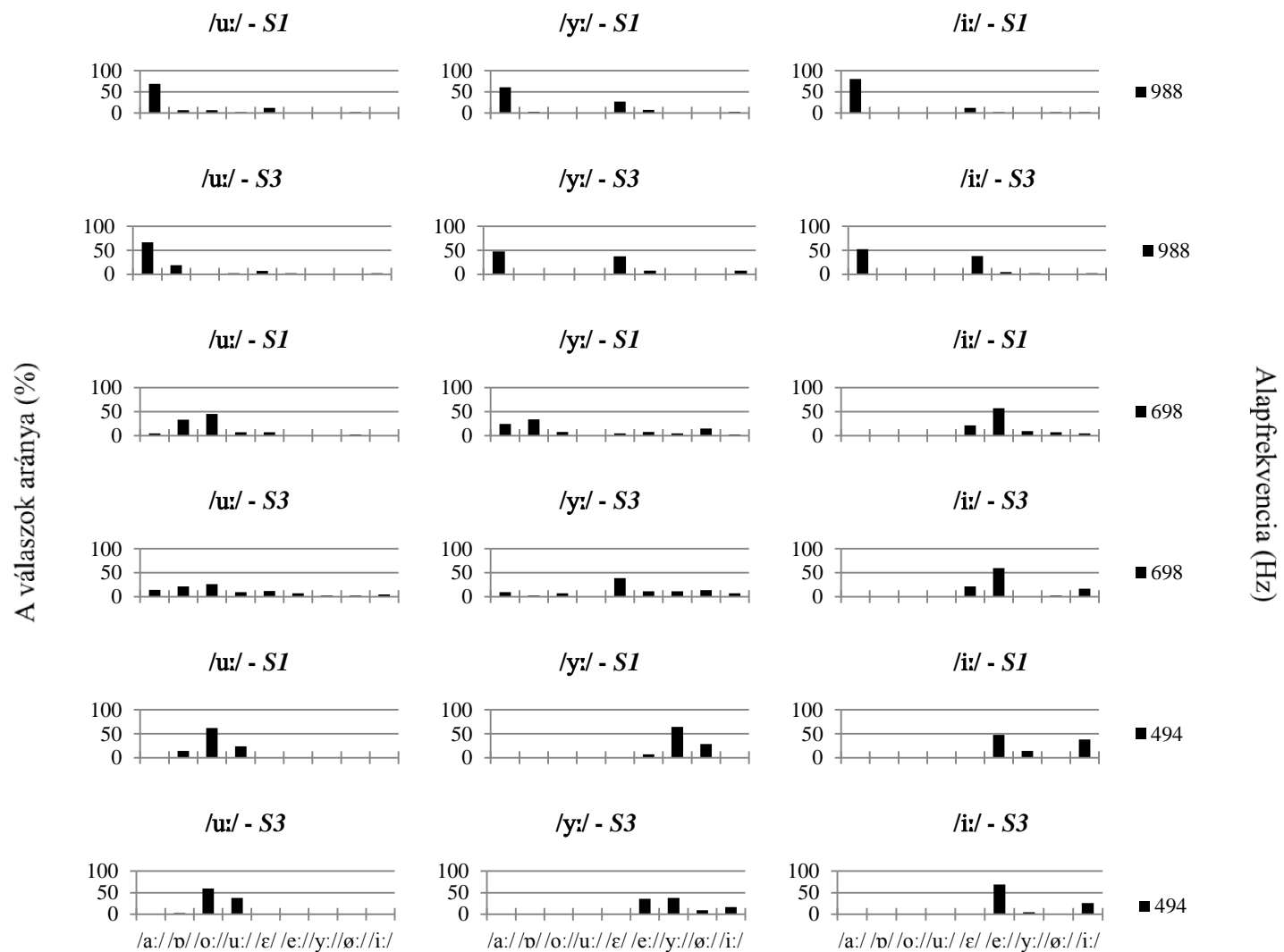
leti csúcs jelenik meg, és e hangok esetében a csökkenő  $F_1$ - $f_0$  távolság zártabb hangzókat eredményez az észleletben. A három második nyitásfokú hangzóból csak az /o:/ és /e:/ esetében lehetséges a magyar magánhangzórendszerben olyan módon nyíltabbként észlelni az adott hangzókat, hogy az csak egy jegyben, mégpedig a nyíltsági fokban eltérő hangzókként való azonosítást jellemlentsen. (Az /ø:/ esetében ugyanis a nyíltabbként azonosításhoz a nyíltabb és ajakrétes, tehát két jegyben eltérő /ε/ magánhangzót kellene azonosítani, vagy pedig a két fokkal nyíltabb és hátul képzett /ɒ/-t). Így a nyíltság észlelésének szempontjából a második nyitásfokú, középső nyelvállásfokú hangok közül csak az /o:/ és /e:/ magánhangzókat vizsgáltuk meg mindhárom énekes ejtésében (38. ábra).

A három énekes ejtette /o:/ és /e:/ magánhangzók azonosítására az első három alaphangfrekvencián (370 Hz alatt) nagyon hasonló tendenciák voltak jellemzőek: bár az ejteni szándékozott beszédhangok nagy arányban (még 349 Hz-en is 60-70%-ban) megtartották az ejtési szándéknak megfelelő hangszínüket, az adatközlők az alaphangfrekvencia emelésével mind az /o:/-t, mind az /e:/-t egyre nagyobb számban azonosították első zártsági fokú (rendre /u:/ és /i:/) beszédhangokként. Ugyanakkor az is megfigyelhető, hogy az S3 énekes /o:/ hangzója „kilóg a sorból”, hiszen azt 349 Hz-en a hallgatók nagyobb százaléka azonosította az ejtési szándék szerint, mint a többi, azonos nyíltsági fokú beszédhangot, valamint téveszteni is inkább egy nyíltabb hangzóra, az /ɒ/-ra tévesztették a zártabb /u:/ helyett.

Mivel a harmadik és negyedik nyíltsági fokú /ε ɒ a:/ esetében tendenciózusan ugyanazt az azonosítási mintázatot láttuk mindhárom énekesnél, ezért ezeket a hangokat csak az S1 énekes ejtésében szemléltetjük egy, az énekes összes ejteni szándékozott magánhangzójának azonosítását az összes alaphangfrekvencián bemutató ábrán (39. ábra). Mivel az /ε ɒ a:/ hangok esetében 370 Hz alatt még nem találtunk  $F_1 : f_0$  hangolási tendenciát, feltehető volt, hogy a hangzók zártságészleletét ezen a tartományon (a középső nyelvállású, második nyíltsági fokú hangokhoz hasonlóan) csak az  $F_1$  és  $f_0$  távolsága befolyásolja. A 36. ábra szerint az /a:/ esetében ez a távolság nem lépte át a 3 Barkos, az /ε/ esetében pedig a 2,5 Barkos kritikus határt. Ugyanakkor az /ɒ/ esetében azt látjuk, hogy az  $F_1$  és  $f_0$  távolsága az itt vizsgált második és harmadik alaphangfrekvencián (azaz a zenei  $f'$  és  $h'$  hangokon) megközelítette, sőt át is lépte a kritikusan megállapított (3 Barkos) értéket. Az észlelési adatokban ezzel együtt azt találtuk, hogy az első három vizsgált alaphangfrekvencián (370 Hz

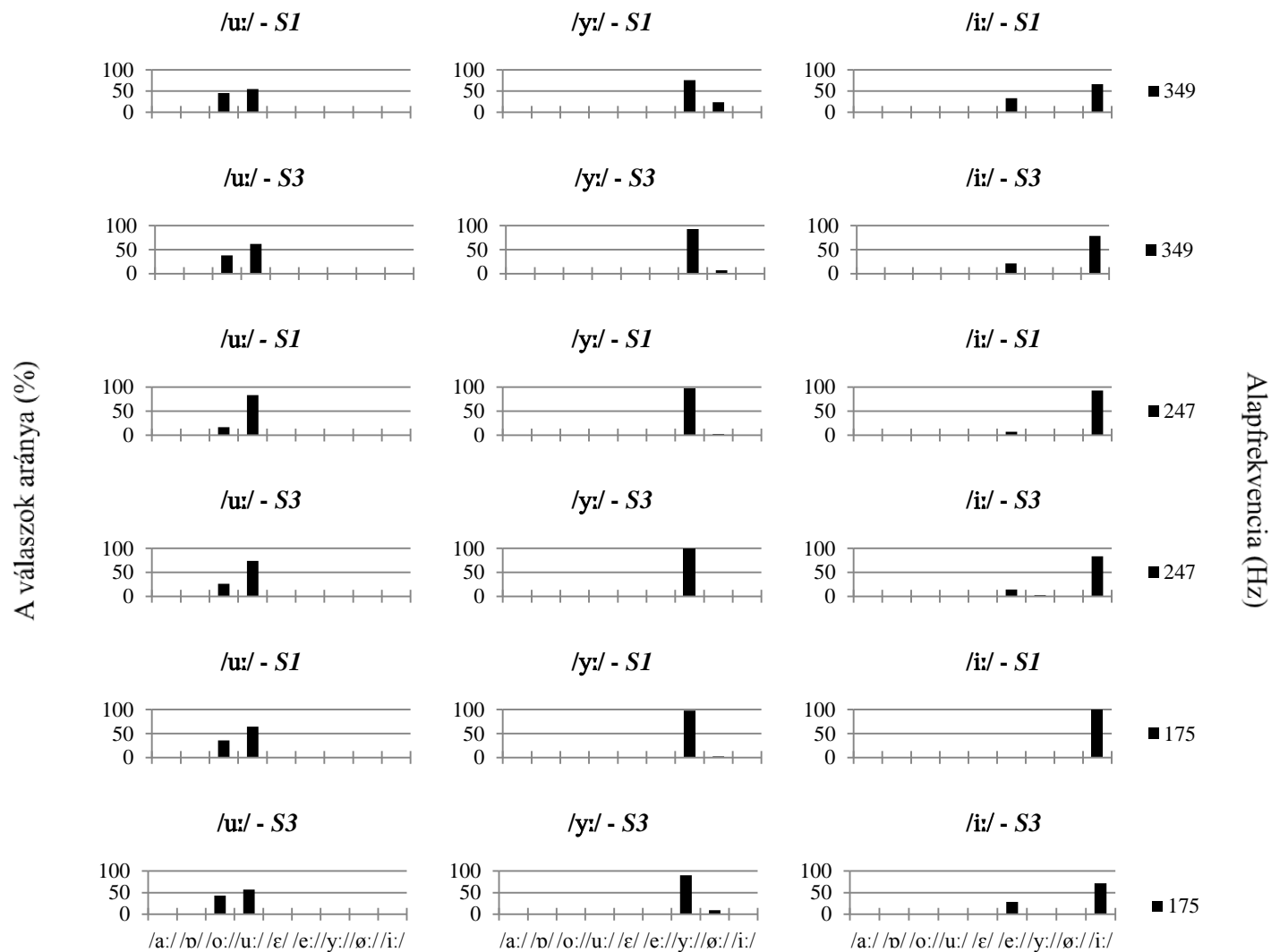
alatt) mindhárom magánhangzót nagy arányban (legalább 70%-ban) az ejtési szándéknak megfelelően azosították az adatközlők. Azonban az is kiderült, hogy az /ɒ/ azonosítása, az alapfrekvencia emelésével a másik két hangnál gyorsabb ütemben csökkent (míg nem 349 Hz-en, a zenei *h'* hangon elérte a 71%-ot). Az /ɒ/-ra az alapfrekvencia emelésével egyre nagyobb számban érkeztek /a:/ válaszok, amit úgy értelmezhetünk, hogy az /ɒ/-t ezek az adatközlők valamivel nyitottabbnak észlelhetnék (hiszen az /a:/ első formánsa magasabb, mint az /ɒ/-é (így észleletileg az /a:/ nyitottabbnak tekinthető, mint az azonos alapfrekvencián ejtett /ɒ/)).

Az integrációs határ átlépése után (ez a jelen kísérletben a zenei *f'* hang, azaz a 349 Hz alapfrekvencia elhagyását jelenti az  $f_0$  emelése során) a tonotópiás modell szerint a beszédpercepció minden ejteni szándékozott magánhangzóban két prominenciát azonosít, mégpedig az  $f_0$  és a  $2f_0$  frekvenciájának magasságában. Az akusztikai adatok alapján láttuk, hogy az első nyíltsági fokú /i: u: y:/ esetében a 349 Hz fölötti alapfrekvenciákon már minden alapfrekvencia és minden énekes esetében fellép az  $F_1$ -nek az  $f_0$ -ra hangolása, tehát az  $f_0$ -nál lévő észleleti prominencia felerősödik. Ha a rezonanciák értéke a beszédben mérhetőkhöz képest állandó maradna, Traunmüller (1981) modellje és kísérletes adatai alapján azt várnánk, hogy a hangzók /i: u: y:/ zártságának észlelete jó arányban (50% fölött) az ejtési szándéknak megfelelő lenne. Az  $F_1$ -nek az  $f_0$ -ra hangolása azonban az  $f_0$ -nál lévő észleleti prominencia erősödésével, tehát az első nyíltsági fokú hangokra jellemző észleleti mintázatnak a bizonyos mértékű megváltozásával jár, a tonotópiás modellből így az /i: u: y:/ azonosítására jellemző tendenciákat nem jósolhatjuk egyértelműen. A 37.a és 37.b ábrán szemléltetett adataink szerint 370 Hz fölött az alapfrekvencia emelésével az ejtési szándék szerint első nyitásfokú, azaz zárt /i: u: y:/ magánhangzókat a hallgatók egyre nyitottabb hangokként, végül pedig a legmagasabb zenei hangon, a *h''*-n (988 Hz alapfrekvencián) legnagyobb arányban /a:/-ként azonosítják: *S1* énekes ejtésében rendre 81%-ban, 69%-ban, valamint 61%-ban, míg az *S3* ejtésében 52%-ban, 67%-ban, valamint 48%-ban. Mindez azt is jelenti, hogy már csak az első nyitásfokú, felső nyelvvállású hangokra adott válaszokként is nagy számban jelennek meg az észleletben középső nyelvvállású hangok is 370 Hz fölött, ami Traunmüller (1981) modelljének a beszéd módra vonatkozó jóslataival nem összeférhető.



37. a. ábra: Az első nyitásfokú, zárt /u: y: i:/ ejteni szándékozott magánhangzókra adott válaszok az f<sub>0</sub> függvényében S1 és S2 énekes ejtésében a 349 Hz-nél magasabb alapfrekvenciákon

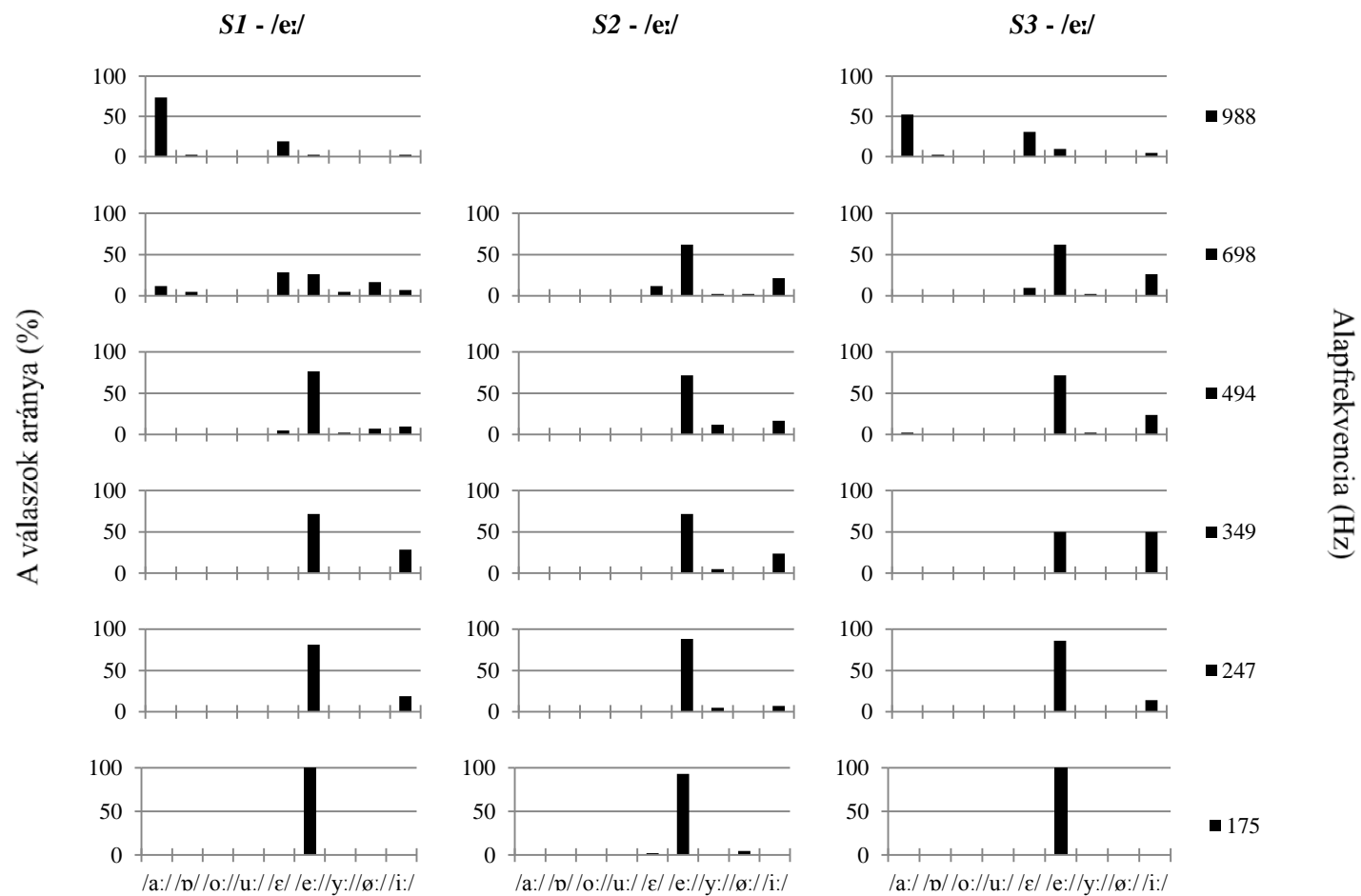




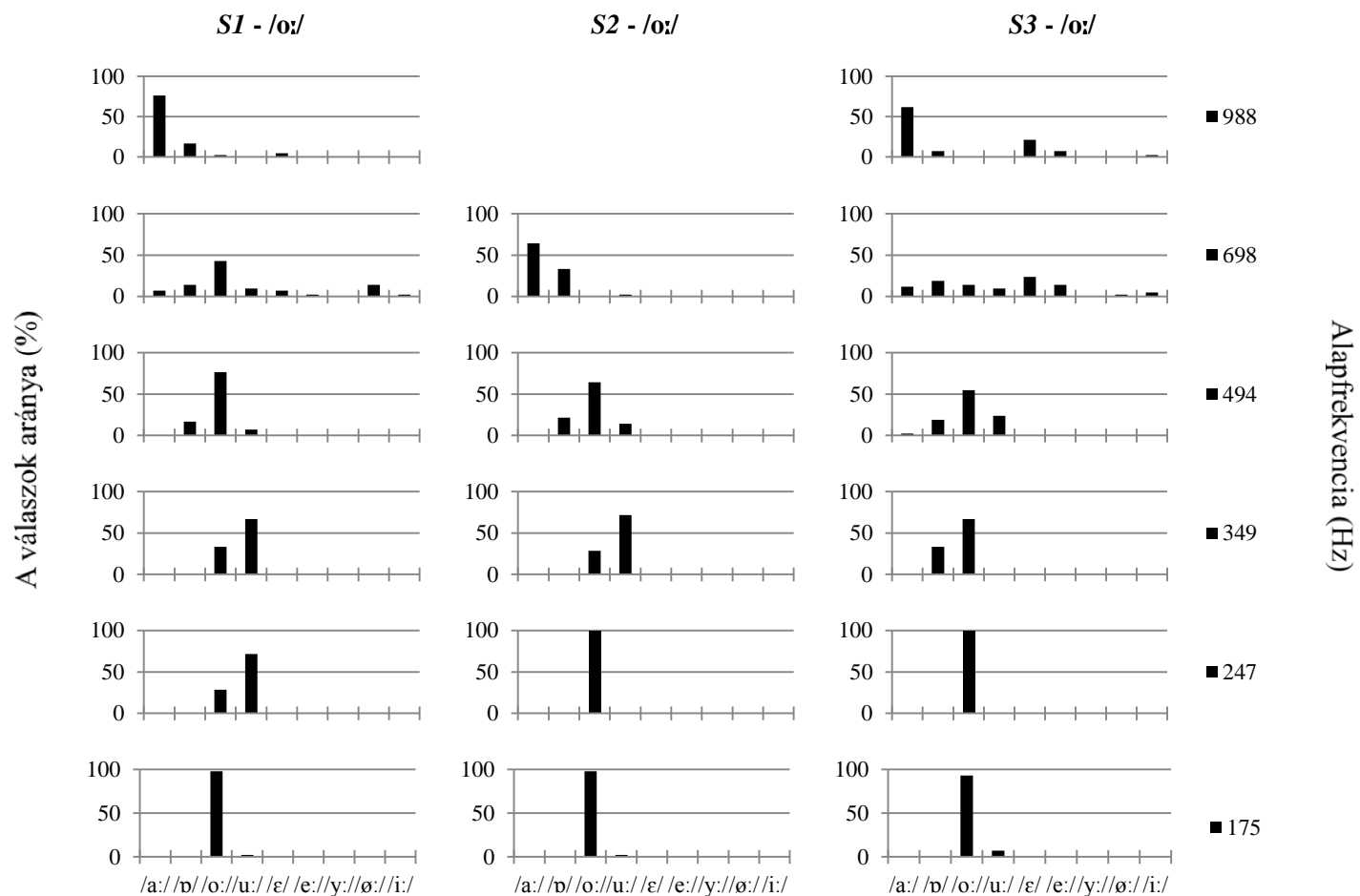
37. b. ábra: Az első nyitásfokú, zárt /u: y: i:/ ejteni szándékozott magánhangzókra adott válaszok az f<sub>0</sub> függvényében S1 és S2 énekes ejtésében a 494 Hz-nél alacsonyabb alapfrekvenciákon

A második nyíltsági fokú /e:/ az  $f''$ -ig (698 Hz-ig) magas, 50% fölötti arányban őrizte meg az ejtési szándék szerinti hangszínét a zártág tekintetében. A 38. ábrán jól látható módon az egyre növekvő számú az ejtési szándéktól eltérő hangminőségű válaszok között a leggyakoribb az első nyitásfokú /i:/ volt, míg a legmagasabb,  $h''$  hang (988 Hz alapfrekvencia) elérésével a válaszok legnagyobb része (az *S1* és *S3* énekesek ejtésében rendre 74%-ban és 52%-ban) az /a:/-ra érkezett. Míg az /o:/-ra a kritikus 370 Hz alapfrekvencia alatti legmagasabb alapfrekvenciáig, az  $f''$ -ig (349 Hz-ig), nagy számban kaptunk /u:/ válaszokat (legalábbis *S1* és *S2* énekes ejtésében), addig a kritikus alapfrekvencia fölött mindhárom énekes ejtésében ismét megnövekedett (és 50% fölé emelkedett) az erre a hangra adott, az ejtési szándéknak megfelelő válaszok száma. A zenei  $f''$  hangon (698 Hz alapfrekvencián) az észlelési tendenciák nem voltak egyöntetűek a három énekes ejtette /o:/ hangokra nézvést. Míg a kísérleti személyek *S1* esetében 48%-ban az ejtési szándék szerint azonosították azokat, *S2*-nél ez az arány jóval magasabb, 64%, *S3*-nál pedig jelentősen alacsonyabb, mindössze 14% volt (ez utóbbi alig valamivel magasabb, mint a véletlen lehetősége, mely ebben a tesztben 1/9 valószínűség, azaz 11% volt). Az /o:/-ra a zenei  $f''$  hangon (698 Hz alapfrekvencián) kapott, a nem az ejtési szándéknak megfelelő válaszok szintén nem voltak egyöntetűek: míg az /o:/-ra *S2* énekes ejtésében a legnagyobb számban /a:/ válaszokat kaptunk (62%-ban), a másik két énekesnél a véletlenhez közeli arányban érkeztek válaszok. A legmagasabb alapfrekvencián, ahogy az /e:/-re is, úgy az /o:/-ra is főként /a:/ válaszokat kaptunk (*S1* és *S3* énekeseknél rendre 76% és 62%-ban).

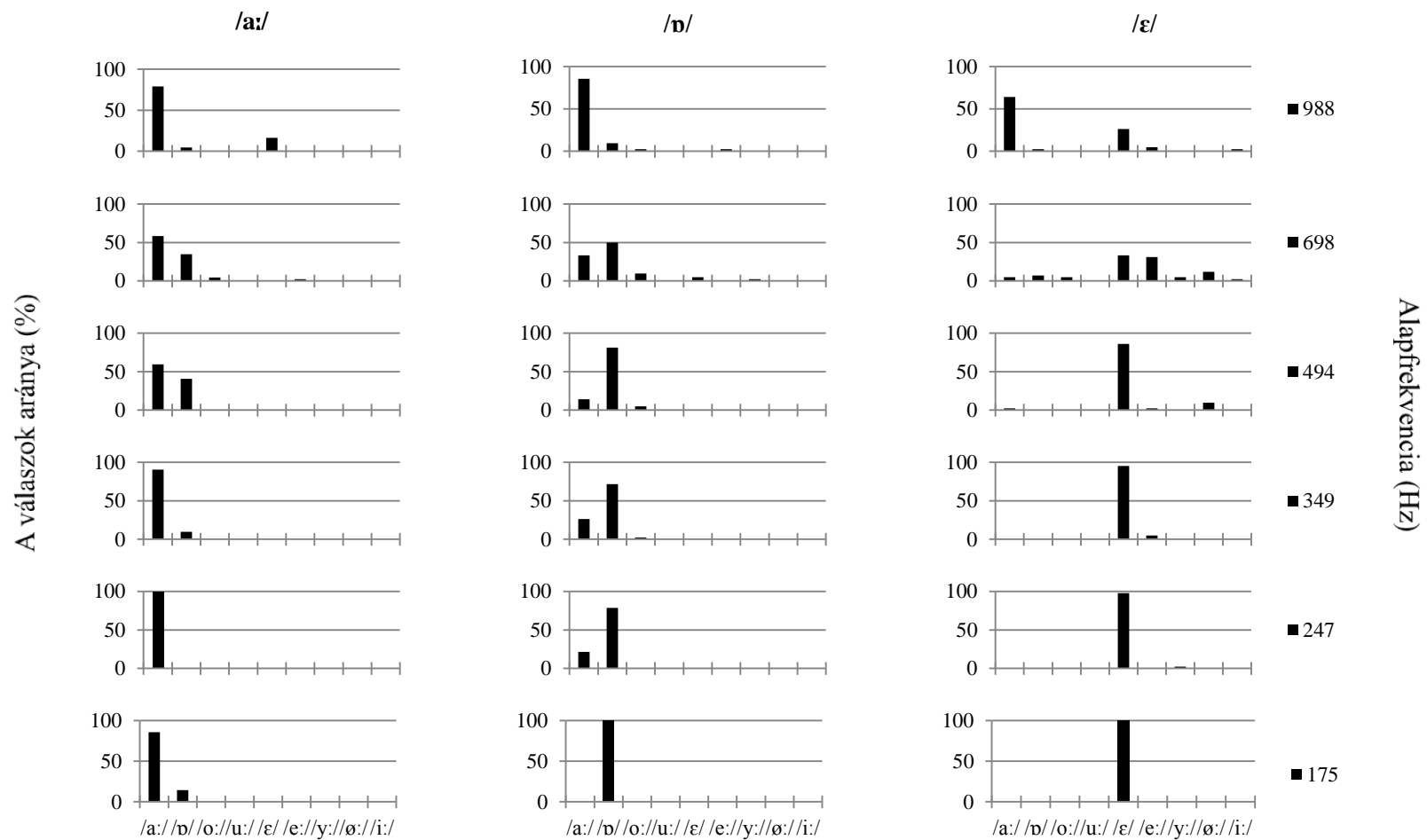
A harmadik és negyedik nyíltsági fokú /ε ɒ a:/ esetében 370 Hz fölött a három énekes ejtette magánhangzókra ismét nagyon hasonló tendenciákat találtunk – így ezeket ismét csak *S1* énekes ejtésében szemléltetjük (a 39.a ábrának az első három sorában található paneleken). A harmadik és negyedik nyíltsági fokú, alsó és legalsó nyelvvállású magánhangzók a zenei  $h'$  hangon (494 Hz alapfrekvencián) még nagymértékben egyértelműen megőrizték érzeti nyíltsági fokukat (az /ε ɒ a:/ rendre 86, 81 és 60%-ban). Ahogyan azonban a többi beszédhangnál is, úgy e hangok esetében is csökkennek az ejtési szándéknak megfelelő azonosítások az  $f''$  alaphangon (698 Hz-en). A legmagasabb, 988 Hz alapfrekvencián azaz a  $h''$  zenei hangon pedig az /ε/ /ɒ/ és /a:/ esetében is a legnagyobb számban /a:/ választ kaptunk (rendre 64, 86 és 79%-ban).



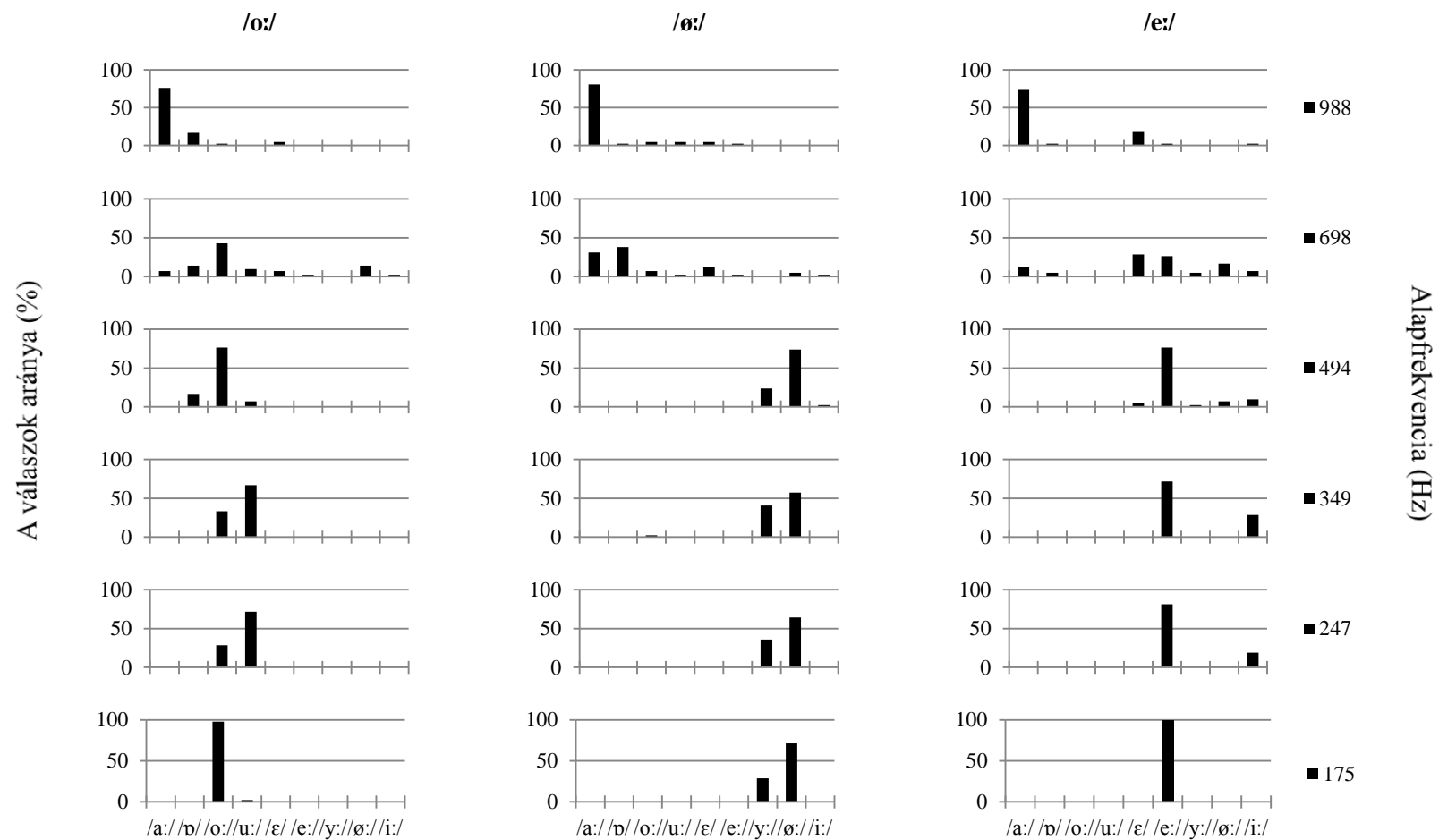
38. a. ábra: A középső nyelvvállású, második nyitásfokú, a nyitásfok szerint csak egy jegyben eltérő szomszédokkal rendelkező /e:/ és /o:/ ejteni szándékozott magánhangzókra adott válaszok az f0 függvényében a három énekes ejtésében



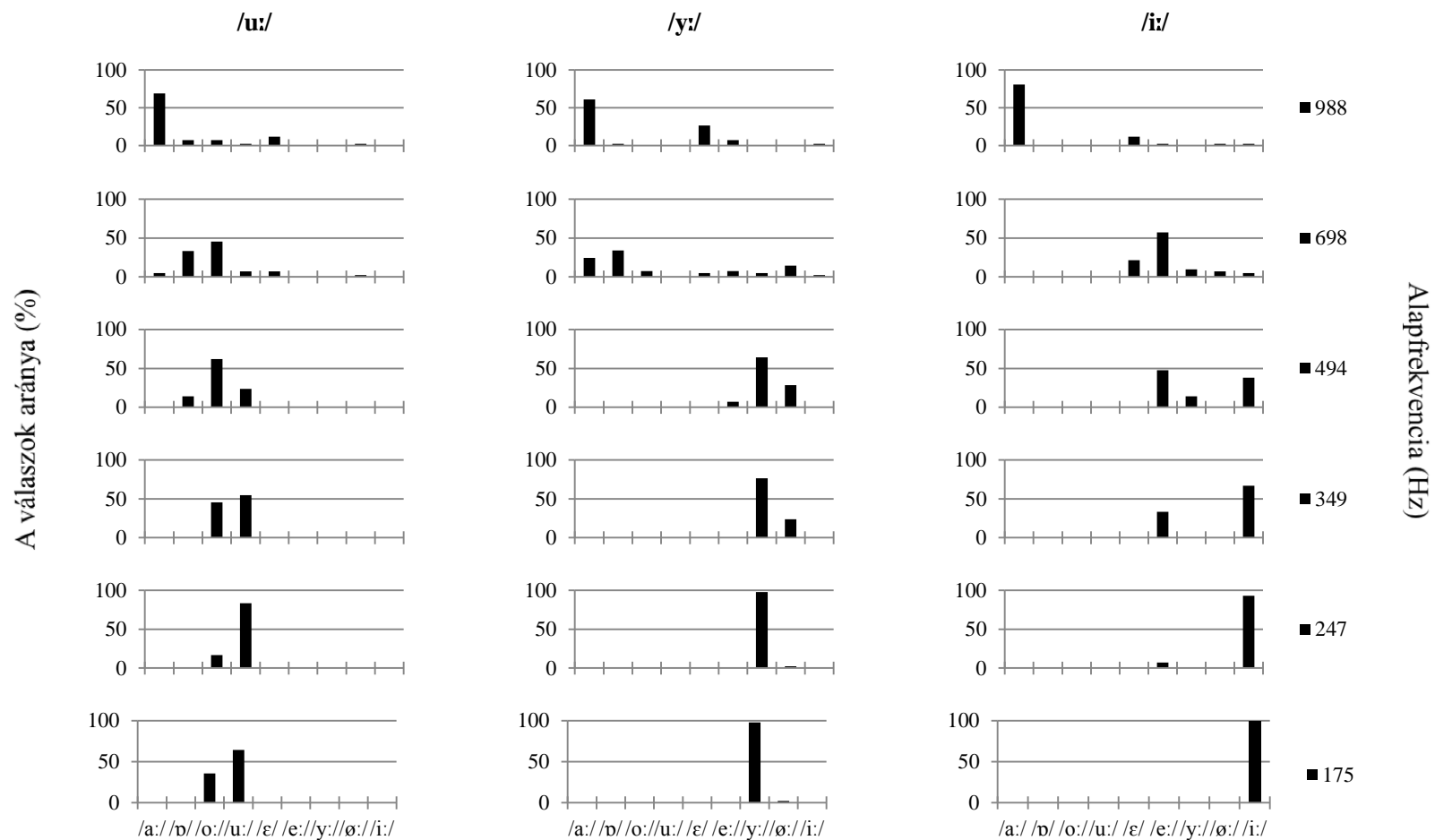
38. b. ábra: A középső nyelvállású, második nyitásfokú, a nyitásfok szerint csak egy jegyben eltérő szomszédokkal rendelkező /e:/ és /o:/ ejteni szándékozott magánhangzókra adott válaszok az f0 függvényében a három énekes ejtésében



39. a. ábra: Az egyes ejteni szándékozott magánhangzókra adott válaszok az alapfrekvencia függvényében *SI* énekes énekprodukciónak

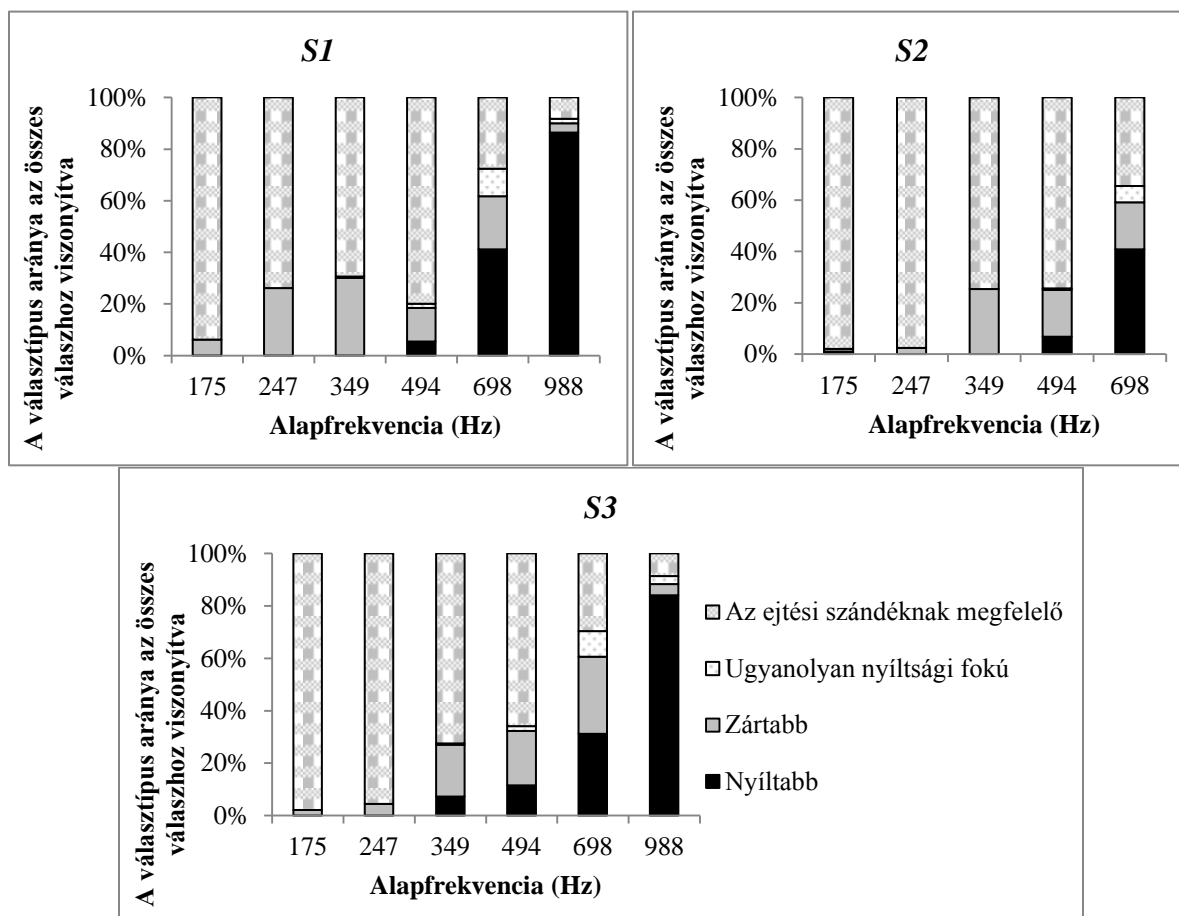


39. b. ábra: Az egyes ejteni szándékozott magánhangzókra adott válaszok az alapfrekvencia függvényében *SI* énekes énekprodukciójában



39. c. ábra: Az egyes ejteni szándékozott magánhangzókra adott válaszok az alapfrekvencia függvényében *S1* énekes énekprodukciónak

Zárásképp ismét az észlelés valamivel általánosabb tendenciáit elemeztük, mégpedig olyan módon, hogy azok összevethetők legyenek a korábbi, a 2.1. és 2.2. *alfejezet*ekben ismertett eredményeikkel is. Kiemeltük a második és harmadik nyíltsági fokú /ε e: ø: o:/ hangzókat (melyek a magyar magánhangzórendszerben nyíltabb és zártabb szomszédokkal is rendelkeznek), és megvizsgáltuk, milyen hangzókként azonosították őket a hallgatók az egyes alapfrekvenciákon az egyes énekesek produkciójában (40. ábra).



40. ábra: A második és harmadik nyíltsági fokú /ε e: ø: o:/ hangzókra az egyes alapfrekvenciákon kapott válaszok megoszlása a magánhangzók nyíltsági fokának függvényében az *S1*, *S2* és *S3* énekesnél

Korábbi vizsgálatainkban 500 Hz és 650 Hz közötti (a temperált zenei hangsorok és a nyugati zene hagyományától független, önkényesen választott) alapfrekvenciákon énekelt magánhangzó-



megvalósulásokat elemeztük, melyekre az ejtési szándéktól eltérő azonosítások esetén dominánsan a hangzók az ejtési szándéknál zártabb minőségű magánhangzókként azonosítása volt jellemző az adott tartományon az alapfrekvenciától függetlenül. A jelen vizsgálatban ezzel megegyezően azt találtuk, hogy a korábban vizsgált alapfrekvenciák környékén –  $h'$  (494 Hz) alaphangon –, illetve az az alatt énekelt magánhangzókra regisztrált észlelési tendenciákat szintén az egyes hangzók zártabb magánhangzókként azonosítása dominálta. Ugyanakkor azt is láttuk, hogy az alapfrekvencia további emelésével (650 Hz fölött), a kifejezetten magas alapfrekvenciákon a hierarchia élére a hangzók az ejtési szándéknál nyíltabb minőségű azonosítása lépett, főként azért, mert – mint azt már említettük – ezeken az alapfrekvenciákon jelentősen megszorodott az /b/ és /a:/ válaszok száma.

#### **2.5.6. Következtetések**

A jelen vizsgálatban szoprán énekeseknek a magánhangzókat érintő formáshangolási stratégiáit vizsgáltuk magas alapfrekvenciájú éneklésben, valamint e magánhangzók azonosításának tendenciáit nem énekes adatközlők észlelésében. Megelőző akusztikai vizsgálatunkkal (2.1. *alfejezet*) ellentétben ez alkalommal korábban leírt módszerek egy új kombinációjának segítségével módunk nyílt nem csak az elhangzó hang spektrális maximumhelyeinek detekciójára, de a produkció közben a rezonátorüreg-rendszerre jellemző sajátfrekvenciák, azaz a formánsok megállapítására is. Ezzel a jelen alfejezetben ismertetett vizsgálat tekinthető az első olyan akusztikai elemzésnek, mely egy adott nyelv összes magánhangzójának az alapfrekvencia emelésével megváltozó produkcióját feltérképezte (közvetlen, akusztikai adatokon keresztül, vö. 1.2. *alfejezet*), különös tekintettel az  $F1 : f_0$  hangolásra. A vizsgálat második részét képező percepcióos teszt eredményeinek elemzésekor az akusztikai adatokból is kiindulva – a jelen értekezés szerzőjének ismeretei szerint – az énekhang vizsgálatában szintén elsőként alkalmaztuk Traunmüller (1981) modelljét, mely a magánhangzó észlelt zártságát az  $F1$  és  $f_0$  tonotópiás távolságából vezeti le.

Magas alapfrekvencián a szoprán operaénekesek gyakran annál a frekvenciasávnál is magasabb alapfrekvenciájú zenei hangok ejtésére kényszerülnek, amelyen a beszélt ejtésben produkált magánhangzók hangszínét jelentősen meghatározó első spektrális maximum, az első formáns realizálódik. Ha viszont az  $f_0$  értéke ilyen módon magasabbra kerül, mint az  $F1$ , a szájat elhagyó

akusztikai jel frekvencia-összetevőinek energiája, tehát az ejtett hang hangossága is jelentősen lecsökken, a hangzó hangszínezete pedig hirtelen és drasztikus (az éneklésben nem kívánatos) változást szenved el. Bár a hangosságélményt érintő nem kívánt hatás megnövelt szubglottális nyomással némileg kompenzálható lenne – azaz hangerő-növekedést idézne elő, ha az énekes a fent említett helyzetben több levegőt és nagyobb intenzitással préselne ki a tüdejéből a hangadáshoz –, a nagyobb szubglottális nyomással fenntartott, huzamosabb fonáció a hangszalagok megerőltetését okozná, tehát hosszú távon káros, és a jó technikájú éneklésben mindenképpen került megoldás. Éppen ezért az énekhang szakirodalmában elterjedt nézet szerint az énekes e helyett a káros technika helyett a megemelt alapfrekvencia hatására fellépő hangerőcsökkenést és hangszínváltozást a formánsok frekvenciáit befolyásoló stratégiákkal kompenzálja. Ez pedig elsősorban az  $F1 : f_0$  hangolás – melynek valós megjelenésére azonban (amint azt az 1.2. alfejezetben részletesen tárgyaltuk) a mérésmódszertan korlátai miatt viszonylag kevés empirikus bizonyítékkal rendelkezünk. Ráadásul az is igaz, hogy a korábbi mérések egyike sem célozta meg egy adott nyelv teljes magánhangzó-készletének vizsgálatát. A formánsstratégiák alkalmazásával, azaz a megváltozott beszédhangejtéssel párhuzamosan természetesen az is kérdéses lesz, miként azonosítják a hallgatók ezeket a magas alapfrekvenciákon énekelt magánhangzókat. Erre a kérdésre, valamint arra, hogy milyen tendenciák jellemzik a nem az ejtési szándék szerinti azonosításokat és miért, eddig nem született kielégítő válasz.

A jelen alfejezetben bemutatott vizsgálat eredményei alapján megfogalmazhatjuk, hogy a kísérletben vizsgált három profi szoprán énekes egységes és általánosan leírható formánsstratégiákat alkalmazott, továbbá az eredmények a korábban, más metodológiával és más nyelvek hangzóira kapott eredményekkel is összhangban vannak. A jelen vizsgálatban igen nagy mennyiségű anyagon, a magyar nyelv összes magánhangzóján, széles (a szopránok hangterjedelmét lefedő három oktávnyi) hangterjedelemben, három énekes produkciójában sikerült demonstrálnunk azt a nemzetközi szakirodalomban még csak néhány nyelv bizonyos hangzóira leírt stratégiát, mely szerint azokban az esetekben, amikor az  $f_0$  megközelíti az adott magánhangzó beszédbeni ejtésére jellemző  $F1$  értékét, az énekes az  $F1$  formánst a megemelt  $f_0$  értékére (vagy valamivel a fölé) emeli. Fontosnak tartjuk hangsúlyozni, hogy bár az adataink az  $F1$  formáns hangolását egyértelműen alátámasztják, a formáns frekvenciaértékének megváltoztatásához használt artikulációs

stratégiákat pusztán a jelen vizsgálat akusztikai adataiból nem következtethetjük ki. Adataink tehát csak és kizárólag a rezonanciák hangolásának akusztikai alapú stratégiáit demonstrálják, nem pedig azokat az artikulációs stratégiákat, melyeket az énekesek ezen akusztikai célok eléréséhez alkalmaznak. Ennek megfelelően tehát abban az éneklés szakirodalmában vitatottnak tűnő kérdésben sem tudunk állást foglalni, mely az F1 hangolását előidéző artikulátorok mibenlétét firtatja a különböző nyelvvállásfokú hangok esetében: nem rendelkezünk ugyanis információval arra nézvést, hogy a kérdéses akusztikai stratégiát az állkapocsnyitás szögének növelése, vagy inkább a nyelvvállásfok csökkentése okozza a zárt ejtésű, felső nyelvvállású magánhangzóknál (vö. Sundberg–Skoog 1995).

Eredményeink egyértelműen azt mutatják, hogy a kísérletben vizsgált magánhangzóknál az  $F1 : f_0$  hangolási tendencia megjelenése az F1 beszédbeni értékétől függött, azaz az F1 hangolásának alsó határa a magánhangzók nyíltsági fokával vagy a képzésükhöz szükséges vízszintes nyelvmozgással és/vagy az állkapocsnyitás szögével volt összefüggésben. A legalacsonyabban (az  $f'$  zenei hangon) a zárt, első nyíltsági fokú vagy felső nyelvvállású magánhangzók F1 formánsát hangolták az énekesek az  $f_0$  értékére, melyet a második nyíltsági fokú, középső nyelvvállású hangok követtek (az  $f''$  zenei hangon), majd pedig a harmadik és negyedik nyíltsági fokú, alsó és legalsó nyelvvállásfokú hangok (a  $h''$  zenei hangon). A jelen vizsgálat eredményei javában egyetértésben vannak Joliveau és munkatársainak (2004) akusztikai eredményeivel, ám az idézett kísérletben talált egyik jellegzetességet saját eredményeinkben nem tapasztaltuk: a jelen kísérletben ugyanis az énekesek nem érték el az  $F1 : f_0$  hangolásának felső határát az ajakkerekítéses magánhangzóknál 900 Hz körül. Megkockáztatható az a feltételezés, miszerint a két vizsgálat közti különbség egy, a formánshangolás mint énektechnikai stratégia és az énekes hangtartománya közti összefüggésre utalhat.

Joliveau és munkatársai (2004) tanulmányának eredményei alapján úgy tűnik, hogy a vizsgálatban részt vett énekesek közül többen elérték az 1042 Hz alapfrekvenciát (vagy akár magasabb alapfrekvenciákat is), tehát feltehető, hogy ezen énekesek hangtartományának felső határa magasabb volt, mint a fentebb említett, a hangolás szempontjából kritikusnak tűnő 900 Hz. Garnier és munkatársai (2010) vizsgálata szerint azonban az  $F1 : f_0$  hangolásának generikus és magánhangzófüggetlen (és inkább az ember anatómiai és fiziológiai működései által meghatározott) felső

határa van, mégpedig 1100 Hz körül. Ettől az alapfrekvenciától kezdve ugyanis az  $F1 : f_0$  hangolás nem lehetséges (mert nem lehetséges az állkapocsszög további nyitása sem). Ezzel összefüggésben a szerzők a vizsgálatban 1100 Hz fölött más stratégiák, például az  $F2 : f_0$  hangolás megjelenését találták – de természetesen csak abban az esetben, ha az 1100 Hz fölötti frekvenciatartomány még az énekes által használt hangtartományához tartozott. A jelen vizsgálat adatközlői – amennyiben a hangtartományt definiálhatjuk a magas hangok viszonylag tartós éneklésére való képesség meglétével – úgy tűnt, megközelítőleg elérték a hangtartományuk felső határát a 988 Hz alapfrekvenciájú  $h''$  hangon. Ha tehát feltesszük, hogy a jelen vizsgálat adatközlőinek hangtartományát így megközelítőleg lefedtük a  $h''$  hangig terjedő alapfrekvenciákkal, és adatainkat a korábbi vizsgálatok fényében értelmezzük, két következtetésre juthatunk.

Az egyik következtetés szerint a jelen vizsgálat adatközlőiként szolgáló énekesek a  $h''$  zenei hang, azaz 988 Hz alapfrekvencián elérték a hangtartományuk (fiziológiai paraméterek által befolyásolt) felső határát, így nem volt szükségük más stratégiák (például az  $F2 : f_0$  hangolás) ismeretére, viszont ez azt is jelentette, hogy képesek voltak kiterjeszteni az  $F1 : f_0$  hangolási stratégiát a hangtartományuk legmagasabb alapfrekvenciáig még az ajakkerekítéses hangzók esetében is. A másik lehetséges értelmezés szerint viszont az is feltételezhető, hogy maguk a formánsstratégiák befolyásolják a hangtartomány kiterjedtségét, tehát az énekesek éppen azért érték el a hangtartományuk felső határát, mert nem tudnak más, az  $F1 : f_0$ -t helyettesítő stratégiákat alkalmazni (például az  $F2 : f_0$  hangolást). Azt, hogy a két felvetés közül melyik helytálló, tehát hogy az énekes által birtokolt formánsstratégiák következnek-e a más faktorok által meghatározott hangtartományból (és annak kiterjedéséből), vagy pedig inkább a formánsstratégiák determinálják-e a hangtartományt, ehelyütt – célzott vizsgálat híján – nem áll módunkban eldönteni, de további vizsgálatokra feltétlenül érdemesnek tartjuk a kérdést.

Az énekelt magánhangzók  $F2$ -értéke a magánhangzók képzéséhez szükséges vízszintes nyelvműködés függvényében változott: az eredmények alapján úgy tűnik, hogy e tekintetben a veláris magánhangzók a palatális hangoknál érzékenyebbek az  $F1$  hangolására (és az azt előidéző artikulációs változásokra). Az eredmények szerint a hátul képzett magánhangzók  $F2$ -értéke az  $F1 : f_0$  hangolás megjelenésétől kezdve növekedni kezdett az alapfrekvencia emelésével. Ezzel szemben az elől képzett hangzóknál csak kisebb mértékű változásokat találtunk, e hangok esetében az  $F2$

értéke valamelyest csökkent a legmagasabb alapfrekvenciákon. A jelenséggel kapcsolatban – Joliveau és munkatársai (2004) magyarázatát is kölcsönözve – azt feltételezzük, hogy az F2-t érintő változás nem szándékos, az F2 közvetlen hangolását célzó stratégia eredménye, hanem az F1 hangolását kísérő törvényszerű következmény. Ahogy ugyanis az énekes növeli az állkapocs nyitásszögét az f0 emelése érdekében, a nyelvhat által okozott szűkület is törvényszerűen egyre nagyobb lesz (a veláris vagy a palatális régióban). Mivel a nyelv és az állkapocs bizonyos mértékig külön-külön is működtethető artikulációs szervek, az állkapocsnyitás növelése miatt egyre kisebb (a nyelvhat és a száypad közötti) szűkület kompenzálható, tehát a szűkület közel eredeti mérete fenntartható a nyelvhat magasságának intenzív növelésével az állkapocs vízszintes helyzetének csökkentése mellett is. Lindblom és Sundberg (1971) egy artikulációs modellen alapuló vizsgálatának eredményei szerint azonban ez a kompenzáció csak egy bizonyos mértékű állkapocsnyitásszög eléréséig lehetséges. Ha ugyanis a metszőfogak távolsága eléri a 23 millimétert, a nyelv már nem képes megfelelően kicsi szűkületet létrehozni – legalábbis nem a veláris területen –, így a magánhangzó centralizálódni kezd (előrébb tolódik az akusztikai magánhangzóterben), illetve második formánsának értéke növekedni fog. Az állkapocs nyitásszögének növelése tehát artikulációs értelemben egy adott határértéken túl törvényszerűen eredményezi az ejtési szándék szerint hátul képzett hangzók centralizálódását is. Lindblom és Sundberg (1971) szerint bár az állkapocsnyitásszög növelése a palatális hangzók ejtésére is hatással van, azok a velárisokhoz képest csak jóval nagyobb állkapocsnyitás esetén kezdenek centralizálódni. Ennek eredményeképpen, ha az állkapocsnyitás szöge nő, a velárisok F2-értéke már viszonylag alacsony állkapocsnyitásfoknál növekedni kezd, ám a palatálisok F2-értéke csak jóval nagyobb nyitásszögnél változik. Így az állkapocsnyitásszög növelése a magánhangzók ejtésekor az elől és hátul képzett magánhangzók akusztikai konvergálását eredményezi, mely predikciót a jelen vizsgálat eredményei is alátámasztják.

Az alapfrekvencia emelését kísérő, az F1-et és F2-t érintő változások „végeredményeként” a legmagasabb alapfrekvencián az F1 az első harmonikussal, tehát az f0-lal, az F2 pedig a második harmonikussal, tehát a 2f0-lal esett egybe. Ez viszont nem jelent mást, mint a magyar magánhangzók eltolódását az akusztikai magánhangzóterben a legnyíltabb és centrális pozíció, azaz az /a:/ magánhangzó pozíciója felé.

A manuális inverz szűrés korábban már eredményesnek bizonyult a formánsmeghatározásban 500 Hz alatti férfi énekhangok esetében (vö. Sundberg et al. 2013) és magas alapfrekvenciájú női énekhangok esetében (legalábbis az /a/ magánhangzónál, vö. Hertegård–Gauffin 1993). A jelen vizsgálatban az inverz szűrést Sundberg és munkatársainak (2013) vizsgálatához hasonlóan Henrich és munkatársainak (2004) eredményeire is alapozva egy elektrolottográf által rögzített elektrolottogram segítségével (még pontosabban az elektrolottogram deriváltjának segítségével) végeztük el. Az inverz szűrés során kapott eredmények megbízhatóságát egy külső gerjesztés (elektromos gége) alkalmazásával is biztosítottuk, mely lehetővé tette az énekesek toldalékcső-rezonanciáinak a zöngétől és a zöngé alapfrekvenciájától független detektálását. A vizsgálati eszközök ezen új kombinációja hatékony módszernek tűnik a formánsmeghatározásban még magas alapfrekvenciákon is. A jelen vizsgálat további újdonsága, hogy először kísérelte meg rendszerint elemezni a magas alapfrekvenciájú éneklést kísérő akusztikai változásokat egy nyelv összes magánhangzójának szisztematikus vizsgálata révén.

Az énekelt magánhangzók azonosítására vonatkozó előzetes feltételezéseink részben teljesültek. Míg 988 Hz alapfrekvencián (a zenei *h''* hangon) az adatközlők legnagyobb számban /a:/ magánhangzókat azonosítottak az ejtési szándéktól függetlenül, alacsonyabb alapfrekvenciákon jóval változatosabb megoszlást tapasztaltunk a válaszokban, míg általánosan a magánhangzók az ejtési szándék szerinti azonosításában fokozatos csökkenést láttunk az alapfrekvencia emelésével. Bár valóban azt figyelhettük meg, hogy a hallgatók nagy arányban azonosítanak zárt hangzókat is a magasabb alapfrekvenciákon, az az elvárásunk nem teljesült, hogy az egyes ejteni szándékozott középső nyelvvállású (második nyitásfokú) beszédhangokat kizárólag vagy legalább dominánsan zártakként ismernék fel a hallgatók. Az általános tendenciák korábbi eredményeinkkel összhangban voltak, beleértve azt az eredményt is, hogy a 650 Hz alapfrekvencia alatti hangok esetében a magánhangzóknak nem az ejtési szándék szerinti azonosítását az ejtési szándéknál zártabb minőségű hangzók azonosítása jellemezte. Korábbi eredményeinket ezen túlmenően kiegészíti az az itt kapott tapasztalat is, hogy az ennél a 650 Hz alapfrekvenciánál magasabb alapfrekvenciákon énekelt magánhangzókra megnövekedett az ejtési szándéknál nyíltabb minőségű hangzóválaszok száma, tehát itt már a Scotto di Carlo és Germain (1985), valamint Hollien és munkatársainak (2000) vizsgálataiban sugalltaknak megfelelő eredményeket kaptunk.

Mivel a beszédhangok azonosításában az F1-nek az akusztikai elemzésekben megfigyelt változása miatt a legfontosabb kérdésnek az ejteni szándékozott hangok nyitásfokának azonosíthatósága tűnik, de a hallási észlelési folyamatok frekvenciaérzékenysége miatt tudjuk, hogy nem lehet egyszerű összefüggéseket vonni a formánsértékek és az észlelet között, észlelési adatainkat a zártságot az F1 és  $f_0$  tonotópiás távolságából levezető elmélet kereteiben (Traunmüller 1981), az érzetileg transzformált akusztikai adatainkat is felhasználva értelmeztük.

Az elméletből kiindulva azt vártuk, hogy a modell szerint kritikus 370 Hz alatt, azaz a kísérletben vizsgált 349 Hz ( $f'$ ) alapfrekvencián és az alatt, az F1 :  $f_0$  hangolás megjelenése előtt a hangzók zártságának észleletét az F1 és  $f_0$  észleleti távolsága határozza meg. Az eredmények megfeleltek a modell predikcióinak: a kérdéses alapfrekvenciákon a tesztelők nagy számban azonosították az ejtési szándék szerint az egyes nyíltsági fokú beszédhangokat akkor, ha az F1 és  $f_0$  távolsága a Traunmüller (1981) kísérleteiben az egyes nyíltsági fokokra jellemzőnek megállapított értékeken belül maradt, ha pedig a távolság a kritikus értékek alá csökkent, a válaszok között – tehát az észleletben – zártabb hangzók is megjelentek. Az F1 :  $f_0$  hangolás megjelenése 370 Hz alatt nem látszott befolyásolni az adatokat – feltehetőleg azért, mert a hangolást csak a legzártabb beszédhangok esetében tapasztaltuk, a hangolás eredményeként pedig az inkább a zárt hangzókra jellemző prominenciamintázathoz hasonló mintázatot feltételezhetünk 370 Hz alapfrekvencia alatt.

A zenei  $h''$ -n, azaz 988 Hz alapfrekvencián talált nagyszámú /a:/ válaszokat az akusztikai adatok ismeretében (bizonyos mértékig) Traunmüller (1981) elméletének keretei közt is magyarázni tudtuk. Az elmélet ugyanis a 370 Hz fölött két (az  $f_0$  és a  $2f_0$  közelébe eső) prominenciával rendelkező hangokról azt jósolja, hogy azok /a/-szerű hangokként ismerhetők fel akkor, ha a második észleleti prominencia intenzív (vagy intenzívebb, mint az első prominencia). Bár a zenei  $h''$  hangon ejtett magánhangzók mindegyikénél F1 :  $f_0$  hangolást tapasztaltunk, tehát a hangok spektrumának fizikai értelemben legintenzívebb komponense az  $f_0$  volt, mégis, az észlelési adatok alapján feltehetjük, hogy ezek a zenei  $h''$  hangon (azaz 988 Hz alapfrekvencián) létrejövő prominenciamintázatok erősen hasonlítottak a negyedik nyitásfokú (/a/-szerű) hangokra. Akusztikai adatainkban azt láttuk, hogy az  $f_0$  fokozatos emelését nem csak az F1 emelése kíséri, hanem az F2 változása is: az F2 ugyanis a velárisok esetében egyre növekvő, a palatálisok esetében pe-

dig enyhén, de csökkenő frekvenciaértéket vett fel. Az F2 változásának „végeredménye” a *h* hangon (988 Hz alapfrekvencián) az volt, hogy az F2 frekvenciaértéke (gyakorlatilag minden ejteni szándékozott beszédhang esetében) a zöngé 2f<sub>0</sub> komponensével esett egybe, tehát erősítette azt. Úgy tűnik tehát, hogy az F1 és F2 egybeesése az f<sub>0</sub> és 2f<sub>0</sub> komponensekkel olyan prominenciamintázatot eredményezett, mely viszonylag egyértelműen meghatározta a tesztben részt vevő hallgatók számára az /a/-szerű hangminőséget – szemben az alacsonyabb frekvenciákon tapasztaltakkal, ahol a 2f<sub>0</sub>-hoz eső prominencia kisebb intenzitású volt (hiszen nem erősítette azt az F2), és az /a/ és /b/ válaszok aránya alacsonyabb.

Az észlelési eredmények 370 Hz alapfrekvencia fölött (mely alapfrekvencia fölött az F1 : f<sub>0</sub> hangolás a második nyíltsági fokú, középső nyelvvállású hangokban is megjelent) és a zenei *h* (988 Hz alapfrekvencia) alatt azonban nem voltak Traunmüller (1981) elméletéből és eredményeiből jó arányban bejósolhatók. Traunmüller (1981) elmélete alapvetően nem ad (vagy legalábbis nem egyértelmű) predikciót a 370 Hz fölötti alapfrekvencián alacsony F1 értékkel ejtett beszédhangok azonosítására, illetve egyáltalán nem tér ki olyan esetekre, melyek (bármilyen más módon) az F1 : f<sub>0</sub> hangolt beszédhangok esetéhez hasonlóak volnának, így az azokkal kapcsolatos jóslatokhoz útmutatásul szolgálhatnának. Traunmüller kísérleti eredményeiből kikövetkeztethető azonban mégis azt feltételezhattuk, hogy a 370 Hz fölötti alapfrekvencián ejtett, alacsony F1 értékű hangzókban az alacsony F1, tehát például az F1 : f<sub>0</sub> hangolás is, az első nyitásfokú, zárt vagy felső nyelvvállású hangok észleleti megőrzését segít(het)i, valamint esetleg az egyes, a nem az ejtési szándék szerint azonosított hangzók zártakként való felismerését is előidézheti. A modell 370 Hz alapfrekvencia fölött mindemellett a középső nyelvvállásfokú hangok csak igen alacsony arányú azonosíthatóságát jósolja (Traunmüller tesztjében 50% alatti válaszarányt mutatott be 13 vetélkedő válaszlehetőség, tehát 1/13, azaz 8% valószínűség esetén).

A jelen vizsgálatban a várakozásokkal ellentétben a középső nyelvvállású hangok a jósoltnál jóval nagyobb arányban jelentek meg (akár az ejtési szándék szerinti, akár attól eltérő) válaszként 370 Hz alapfrekvencia fölött. Ezen túlmenően pedig nem tapasztaltuk azt, hogy a hangzókat nagy számban zártakként, azaz első nyíltsági fokúakként azonosították volna a hallgatók, mert azokban az esetekben, amikor az F1 : f<sub>0</sub> hangolás megjelent, egyaránt láttunk zártabb és nyíltabb hangzókként azonosítást is. A vizsgálatunk tehát azt demonstrálta, hogy a hallgatók az operaéneklés-



ben 370 Hz alapfrekvencia fölött is azonosíthatnak harmadik nyitásfokú hangzókat – szemben a beszéddel, ahol elméletileg a középső nyelvállású hangzók azonosítása (vagy azonosíthatósága) elmarad a zártabb és nyíltabb hangzókétól. Ennek okaként pedig éppen az  $F1 : f_0$  hangolás megjelenését feltételezhetjük.

A korábbi, az énekelt magánhangzók észlelését vizsgáló munkák azt találták, hogy az egyre magasabb alapfrekvencián produkciós tekintetben egyre nyíltabban ejtett beszédhangokat az észlelésben is többé-kevésbé egyre nyíltabbként, az ejtési szándék szerinti hangminőségnél magasabb  $F1$ -gyel rendelkező hangokként azonosítják a hallgatók (Scotto di Carlo–Germain 1984, Benolken–Swanson 1990, Hollien et al. 2000). Bár az idézett szerzők maguk is találtak arra utaló eredményeket, hogy ezek az észlelést és az akusztikumot ilyen egyszerű, lineáris összefüggésbe hozó értelmezések nem szolgálnak a magas alapfrekvenciákon ejtett/énekelt hangzók azonosítására tökéletes magyarázattal, tanulmányaikban nem szentelnek a problémának kifejezett figyelmet. Mivel a nagyon általános tendenciákat, de főként a nagyon magas (1000 Hz körüli) alapfrekvenciákon tapasztalt nagyszámú /a/ válaszokat ez a lineáris megfeleltetés is magyarázza (vö. még a jelen alfejezet vizsgálati eredményeit a második és harmadik nyíltsági fokú magánhangzókat jellemző tendenciákat), korábban nem született az énekelt magánhangzókról olyan elemzés, mely az alaphártya frekvenciaérzékenységet vagy a hallás frekvenciafelbontásának más (frekvenciafüggő) sajátosságait is figyelembe véve igyekezett volna magyarázni az ezzel a produkció és percepció közti lineáris megfelelés feltételezésével nem magyarázható percepciós jelenségeket.

A jelen vizsgálat eredményei szerint a 370 Hz alapfrekvencia alatt és az 1000 Hz közeli alapfrekvencián énekelt magánhangzók észlelésének tendenciái jól magyarázhatók Traunmüllernek (1981) a zártságészleletet az  $F1$  és  $f_0$  tonotópiás távolságából levezető modellje segítségével. A modell keretei között értelmezve ugyanis belátható, hogyan lehetséges 370 Hz alatt a zárt, felső nyelvállású magánhangzók jó arányú felismerése – még az  $F1$  hangolásának, azaz az  $F1$  emelésének fennállása esetén is –, valamint a nyíltabb hangzók zártabbakként való felismerése, ahogy indokoltnak tűnik az is, hogy az 1000 Hz körüli alapfrekvencián énekelt hangokat miért leginkább /a/-szerű hangoknak feleltetik meg a hallgatók. Ugyanakkor az  $F1 : f_0$  hangolás miatt úgy tűnik, 370 Hz és 1000 Hz között a modell nem bír jó magyarázó erővel az énekelt magánhangzók azonosításának értelmezésében. Azt tapasztaltuk ugyanis, hogy az  $F1 : f_0$  hangolás miatt 370 Hz

fölött is megmaradnak az észleletben a középső nyelvállású hangok – melyeknek Traunmüller modellje szerint 370 Hz alapfrekvencia (és az a fölötti tartomány) a beszédben legalábbis nem kedvez. Ezeket, a 370 Hz fölött és 1000 Hz alatt tapasztalt percepció tendenciákat ugyanakkor viszonylag jól magyarázza az akusztikum és percepció lineáris összefüggéseit feltételező értelmezés, mely szerint tehát az egyre nyitottabb ejtésű magánhangzókat, egyre nyitottabb beszédhangokként azonosítják a hallgatók. Ezen tendenciák értelmezéséhez egy korábbi percepció vizsgálat adatai további érdekes adalékkal szolgálnak.

Egy 1959-ben publikált kísérletében Fant a magánhangzók azonosításához szükséges minimális akusztikai információt tesztelte (valamint áttételesen az alapfrekvencia hatását is a magánhangzószerű hangok azonosításában). A tesztben Fant nem komplex hangokat használt, a vizsgálat anyagát különböző periódusidejű, azaz különböző hangmagasságú egyszerű szinuszok képezték; a tesztben részt vevők feladata az volt, hogy az elhangzó szinuszhangokat azonosítsák valamely svéd magánhangzóként. A tesztet Fant háromszor futtatta le, minden alkalommal (körülbelül) 20 svéd anyanyelvű adatközlővel. Bár az eredmények szerint egy-egy szinuszhang több eltérő minőségű magánhangzóválaszt is kiválthatott (míg mások egy magánhangzó-minőségnek sem voltak megfeleltethetőek), az egyes beszédhangok mégis nagyon erős összefüggéseket mutattak bizonyos frekvenciákkal. Fant összefoglalása szerint az egyes hangokra a 8. táblázatban feltüntetett frekvenciák voltak elsődlegesen jellemzőek – az adott válaszok tehát ezeken a frekvenciákon voltak a legnagyobb sűrűségűek

8. táblázat: Az egyes magánhangzó-minőségekre jellemző frekvenciák egy egyszerű szinuszhangokkal való megfeleltetési feladat szerint (Fant 1959)

Nyíltsági fok	Magánhangzó (IPA)	Frekvencia (Hz)
1.	/u/	400
2.	/o:/	625
4.	/ɑ:/	1000
4.	/æ:/	1250
2.	/e:/	2500 (másodsorban 1600 és 3500)
1.	/i:/	4000
1.	/y:/	2000
1.	/ɤ:/	1750 (másodsorban 250)
2.	/ø:/	1750 (másodsorban 250)

Az eredményeket a magánhangzók spektrumával összevetve Fant azt találta, hogy több beszédhang, de kifejezetten az első és második nyíltsági fokú veláris magánhangzók (azaz az /u/ és /o:/) esetében a szinuszokra adott válaszok a beszédhang spektrumára emlékeztető eloszlást mutattak, azaz a legtöbb válasz a spektrum maximumainak környékére esett. Ez az eredmény persze nem meglepő, ha belegondolunk, hogy a velárisok első két formánsa igen közel van egymáshoz, akár könnyen 3 Bark távolságon belülre is eshet – elsősorban a nyitottabb ejtésű velárisok esetében. Így azt mondhatjuk, hogy a velárisok esetében az első két rezonancia frekvenciasávja mindenképpen dominálja az észleletet, vagy a rezonanciák érzeti integrációja következtében létrejövő egyetlen intenzív érzeti prominencia, vagy pedig két, egymáshoz közel eső, így mindenképpen nagy intenzitású prominencia révén.

Fant (1959) vizsgálata arra mutat, hogy a velárisok viszonylag egyértelműen leképezhetők egyetlen (ráadásul 1000 Hz alatti) frekvenciával, mely frekvencia tehát mind a hátulképzettséget, mind pedig a nyelvállásfokot is kulcsolja (auditive). Az eredmények – másként megragadva – tehát azt mutatják, hogy bizonyos (alap)frekvenciák bizonyos hangok észleletének kedveznek. Mindezzel összhangban Fant (1959) kísérletének konkrét elrendezésétől kicsit jobban elvonatkoztatva két feltételezést is megfogalmazhatunk. Egyfelől feltehető, hogy azokban a komplex hangokban, melyekben a Fant kísérletében talált frekvenciák dominánsak, valószínűsíthető lesz egy adott nyelvállásfokú veláris magánhangzó mint magánhangzóválasz az észleletben. Másfelől pedig feltehető, hogy egy másik olyan komplex hang, amelyben emellett a domináns frekvencia mellett egy magasabb frekvenciájú második prominencia is megjelenik, viszonylag egyértelműen azonosítható lesz a velárisal megegyező nyelvállásfokú, de palatális magánhangzóként. Ilyen értelemben tehát azt mondhatjuk, hogy a Fant (1959) által bemutatott 1000 Hz-nél alacsonyabb frekvenciájú, kitüntetett szerepű szinuszok szoros összefüggést mutatnak az észlelt zártsággal.

A jelen alfejezetben bemutatott vizsgálat percepciós eredményeire pillantva feltűnő az egybeesés Fant (1959) eredményeivel – legalábbis az előbbieken megfogalmazott feltételezések értelmében. Míg a jelen vizsgálatban az első nyíltsági fokú /i: u: y:/ hangokat 400 és 600 Hz között kezdték dominánsan egy fokkal nyitottabbakként azonosítani a hallgatók – Fantnál a 625 Hz-es szinuszhang a második nyíltsági fokú hangok megjelenését indukálta. Ugyanakkor a jelen kísérletben a második nyíltsági fokú /o: e:/ hangokat még 698 Hz alapfrekvencián is nagy arányban

azonosították az ejtés szerinti nyíltsági foknak megfelelően a tesztelők, mígnem 988 Hz alapfrekvencián tapasztaltunk radikális váltást az /a:/ válaszok megszorodása miatt – ahogy Fantnál is azt láttuk, hogy a negyedik nyíltsági fokú veláris /a:/-ra az 1000 Hz alapfrekvenciájú szinusz volt a leginkább hasonlatos.

A jelen vizsgálatban ugyanakkor azt is tapasztaltuk, hogy a veláris hangok némiképp eltérő viselkedést mutattak a palatálisokhoz képest, annak megfelelően, hogy az előbbieken az F1 és F2 közelebb, míg az utóbbiakban távolabb helyezkedik el egymástól. Azt láttuk, hogy a veláris /u:/ esetében, ahol az F2 már az  $F1 : f_0$  hangolás megjelenésének alapfrekvenciájától fogva követni kezdi a  $2f_0$ -t, az ejtési szándék szerinti /u:/ már ettől a 349 Hz-es (illetve 494 Hz-es) alapfrekvenciától, azaz az  $f'-h'$  hangoktól kezdve nyitottabbként, előbb /o:/-ként majd /ɒ/-ként és /a:/-ként jelenik meg az észleletben. Ugyanakkor a palatális /i:/ y:/ esetében az F2 csak 698 Hz-en ( $f''$ ) közelítette meg a  $2f_0$ -t, és csak 988 Hz-en ( $h''$ ) esett azzal egybe, aminek megfelelően a palatálisokban a  $2f_0$ -hoz eső érzeti prominencia nem volt olyan intenzív, hogy az ejtett hangzót a  $2f_0$ -nál jellemzően intenzív /a:/-hoz tegye hasonlatossá. A palatálisokra jellemző prominenciamintázat Traunmüller (1981) modelljének megfelelően azt is jelentené, hogy a zártság észlelete megmaradt az /i:/ y:/ esetében a 698 Hz ( $f''$ ) alatti legmagasabb alapfrekvenciáig (azaz a  $h'$ -ig, 494 Hz-ig). Az eredmények pedig tendenciaszerűen igazolják ezt a predikciót, ugyanis azt láttuk, hogy az /i:/ és /y:/ az /u:/-nál magasabb alapfrekvenciáig megtartja a zártság észleletét (az előbbieken, azaz a palatálisok esetében a zártság észlelete 494 Hz-ig marad fenn, míg a veláris /u:/ esetében inkább csak 394 Hz-ig).

Az itt taglalt következtetéseket szem előtt tartva megfogalmazható az a megállapítás, hogy bár az  $F1 : f_0$  hangolás egy-egy adott (magas) alapfrekvencián uniformizálja a beszédhangokat (az észleleti prominencia minden magánhangzóban megegyező frekvenciája révén), mégis, a hangolás egyúttal azt is lehetővé teszi, hogy ezeken a magas alapfrekvenciákon többféle hangminőség, többféle nyíltsági fok is megmaradhasson az észleletben (úgy, hogy azok nem feltétlenül egyeznek az ejtési szándéknak megfelelő beszédhangok nyíltsági fokával). Ilyen értelemben akár azok a konklúziók is megkérdőjelezhetők, melyek a magánhangzó-minőségek szempontjából egyértelműen „vesztésnek” tekintik az alapfrekvencia emelését kísérő formánshangolási stratégiát (l. pl. Joliveau et al. 2004 és Garnier et al. 2010). Ha ugyanis nem lépne fel az  $F1 : f_0$  hangolás, ak-

kor valószínűsíthetjük, hogy a szopránok hangtartományának javán (370 Hz fölött) csak jóval kevesebb (csak az első és negyedik nyíltsági fokba tartozó) hangzókat volna lehetséges nagy számban azonosítani, ami mindenképpen kisebb auditív változatosságot, és így talán kisebb esztétikai élvezetet is jelentene a szopránéneklés hallgatói számára.

### 3. Összegzés és következtetések

A jelen értekezésben bemutatott vizsgálatok célja a magyar magánhangzók akusztikai és percpiciós jellemzőinek leírása volt a magas alapfrekvenciájú éneklésben. Kísérleteinkkel egyrészt az énekelt magánhangzók produkcióját érintő kérdésekre kerestük a választ, másrészt pedig azt kívántuk felmérni, miként dolgozódnak fel az alapfrekvencia függvényében a beszédbeli megvalósulásuktól kisebb vagy nagyobb mértékben eltérő énekelt magánhangzók a beszédpercpiciós rendszerben.

#### 3.1. A hipotézisek vizsgálata

**H1:** *Az énekesek a magas alapfrekvenciákon énekelt magyar magánhangzók első formánsát a megemelt  $f_0$  értékére (vagy valamivel a fölé) hangolják akkor, ha egyébként az  $f_0$  értéke magasabb lenne, mint a magánhangzóra a beszédben jellemző első formáns értéke. Ennek értelmében az  $F1 : f_0$  megjelenése az éneklésben a magánhangzók nyíltsági fokának (illetve beszédben jellemző  $F1$  értékének) függvénye.*

Első hipotézisünket a jelen értekezésben bemutatott kísérletek egyértelműen igazolták – egyetértésben az énekelt magánhangzók akusztikai szerkezetét vizsgáló nemzetközi szakirodalom eredményeivel (vö. 1.2. alfejezet). Akusztikai vizsgálataink során azt találtuk, hogy a szoprán énekesek a megemelt, magas alapfrekvencia értékére emelik a magánhangzók kialakításában részt vevő első üregi rezonancia, azaz az első formáns értékét akkor, ha az alapfrekvencia megközelíti az első formáns beszédbeli megjelenésének frekvenciáját. Az  $F1 : f_0$  hangolás az egyes magánhangzók nyíltsági fokának, illetve első formánsának függvényében jelent meg az alapfrekvencia emelésével, a zárt, első nyíltsági fokú, felső nyelvvállású /i: y: u:/ hangzóknál az  $f'$  zenei hangon (349 Hz), a nyíltabb, második nyíltsági fokú, vagy középső nyelvvállású /e: ø: o:/ hangzók esetében az  $f''$  zenei hangon (698 Hz), míg a legnyíltabb, harmadik és negyedik nyíltsági fokú /ε

n a:/ esetében a *h''* zenei hangon (988 Hz). Ezzel összefüggésben a legmagasabb alapfrekvencián a legnagyobb akusztikai változásokat a zárt, vagy első nyíltsági fokú hangzók, míg a legkisebbeket a legnyíltabb hangzók szenvedték el (2.5. *alfejezet*).

Az első formánsnak az alapfrekvenciától függő változásai mellett bizonyos mértékben a második formáns frekvenciájának szintén alapfrekvencia-függő változását is megállapíthattuk. Az alapfrekvencia emelésével ugyanis az F2 a palatális hangzók esetében enyhe csökkenést, míg veláris hangzók esetében erős növekedést mutatott. A változások eredményeként a vizsgálatainkban előfordult legmagasabb 988 Hz alapfrekvenciájú, azaz az *f''* zenei alaphangon az összes ejteni szándékozott magánhangzó helyett egyetlen uniformizálódott F1 és F2 értékkel rendelkező magánhangzószerű hang valósult meg, melynek prominens frekvencia-összetevői az  $f_0$  és a  $2f_0$  frekvenciaértékén jelentek meg. E hangzó minősége a magyar magánhangzók rendszerét tekintve leginkább az /a:/ beszédbeli megvalósulásához hasonló (2.5. *alfejezet*).

Az első formáns megemelésének célja, hogy az énekes elkerülje azokat a helyzeteket, melyek során – a magas alapfrekvenciák éneklésekor – az alapfrekvencia magasabb lenne, mint a toldalécső első rezonanciája. Ez ugyanis mind hangerősségben, mind hangszínezetben, mind pedig az adott magánhangzó minősége szempontjából nemkívánatos hatásokat eredményez: a hangerősség csökken, a hangszínezetben hirtelen változás áll be, a hangszín, azaz a magánhangzó minősége pedig megváltozik, hiszen a magánhangzóspektrum elveszíti a magánhangzó-minőség kialakítása szempontjából fontosnak tartott első prominenciáját (vö. pl. Sundberg 1979). Bár a hangosság csökkenése megnövelt szubglottális nyomással némileg kompenzálható lenne – más szóval hangerő-növekedést idézne elő, ha az énekes a fent említett helyzetben több levegőt és nagyobb intenzitással préselne ki a tüdejéből a hangadáshoz –, a nagyobb szubglottális nyomással fenntartott, huzamosabb fonáció a hangszalagok megerőltetését okozná, tehát hosszú távon káros, és a jó technikájú éneklésben mindenképpen került megoldás. Éppen ezért az énekhang szakirodalmában mára elterjedtnek tekinthető nézet szerint az énekesek e helyett a káros technika helyett a megemelt alapfrekvencia hatására fellépő hangerőcsökkenést és hangszínváltozást a formánsok frekvenciáit befolyásoló stratégiákkal kompenzálják. Az  $F1 : f_0$  hangolás eredményeként a legalacsonyabb spektrális összetevő, azaz az alaphang vagy alapfrekvencia erősítést nyer, amely egyrészt az egész hangzóspektrum energiájának növekedését, azaz hangerő-növekedést jelent, másrészt

pedig lehetővé teszi a hangszínezet jobb kontrollálását is (vö. Sundberg 1979, 1987). Bár az  $F1 : f_0$  hangolás e pozitív következményei teoretikusan könnyen beláthatók, az  $F1 : f_0$  hangolásra a magas alapfrekvencia okozta spektrális alul-mintavételezettség miatt kevés empirikus bizonyítékkal rendelkezünk. A spektrális alul-mintavételezettség következtében ugyanis a toldalékcső rezonanciái, azaz a formánsok nem állapíthatók meg megbízhatóan a magas alapfrekvenciájú akusztikai jel (a hangnyomás-idő függvény) Fourier-transzformáción alapuló elemzése segítségével.

Az eddigi vizsgálatok, melyek az  $F1 : f_0$  hangolás megjelenését elemezték, vagy artikulációs adatokból, vagy artikulációs adatokból (akusztikai modelleken keresztül) derivált akusztikai adatokból indultak ki, illetve olyan alternatív felvételi és elemzési módszertant használtak, melyek valamilyen módon lehetővé tették a toldalékcső rezonanciáinak az alapfrekvenciától független detektálását (vö. 1.2.2. *alfejezet*). A jelen értekezésben a formánsfrekvenciák meghatározásához ez utóbbi megoldást választottuk: több, korábban a beszéd- és énekhang elemzésében már sikerrel használt alternatív eszköz kombinációját használtuk a formánsfrekvenciák meghatározásához a magas alapfrekvenciájú énekelt magánhangzóknak. Rezonancia-adatainkat manuális inverz szűrés segítségével nyertük (Granqvist é. n.), melyhez a hangszalagok működését rögzítő elektrolottográfias regisztrátumokat is használtunk Henrich és munkatársai (2004) eredményeire is alapozva Hertegård és Gauffin (1993), illetve Sundberg és munkatársai (2011, 2013) nyomán. Emellett pedig „újra-mintavételeztük” a toldalékcsővet egy, a gégénél alkalmazott külső gerjesztés, egy elektromos gége segítségével Sundberg (1975) módszerének alapján, és az így létrejött akusztikai jelet is elemeztük. A kétféle elemzés egymást kiegészítő és validáló módon szerepelt a vizsgálatban.

***H2:*** Mivel magas alapfrekvenciákon az énekelt magánhangzóknak a frekvenciaszerkezete és a spektrális felbontása is megváltozik, a magas alapfrekvencián énekelt magyar magánhangzóknak az ejtési szándék szerinti azonosítása az alapfrekvencia emelésével törvényszerűen csökken. Az  $f_0$  és  $F1$  csökkenő távolsága miatt a magas alapfrekvenciákon ejtett magánhangzók akusztikai megjelenése az ejtési szándéktól eltérő azonosításokban inkább a zárt hangzók felismerését facilitálja.



A második hipotézis csak részben igazolódott. Egyfelől igazoltnak tekinthetjük a magánhangzók ejtési szándék szerinti azonosítási arányának csökkenését az alapfrekvencia emelésével. Eredményeink szerint ugyanis az alapfrekvencia emelésével változó akusztikai szerkezetű magánhangzók egyre kevésbé voltak felismerhetők az ejtési szándék szerint, ráadásul olyan módon, hogy az akusztikai változás a legmagasabb vizsgált alapfrekvenciáig (körülbelül 1000 Hz-ig) inkább fokozatosan, nem pedig kategorikusan változó percepciós minőségeket eredményezett. Ennek eredményeként az egyes ejteni szándékozott magánhangzókra az alapfrekvencia emelésével egyre változatosabb magánhangzóválaszok érkeznek, azaz a válaszok egyre nagyobb mértékben szórtak a magánhangzó-minőségek mentén (2.1., 2.2., 2.4. és 2.5. *alfejezet*).

Másfelől azonban kísérleteink nem igazolták azt a feltevést, hogy a szoprán alapfrekvencia-tartomány magasabb régióiban egyértelműen a hangzók zártabbként azonosítása volna jellemző. Eredményeink szerint 370 Hz alatt (ami nagyjából a zenei  $f\#'$  hangnak felel meg) az elvárásoknak (és Traunmüllernek a perceptuális nyíltságot az  $F1$  és az  $f0$  távolságából levezető 1981-es modellje alapján bejósoltaknak) megfelelően az egyes énekelt magánhangzókat a nem az ejtési szándék szerinti azonosítások esetén alapvetően zártabbakként azonosították a hallgatók. Ugyanakkor 370 Hz ( $f\#'$ ) fölött ez a tendencia megváltozni látszott. Egyfelől ugyanis azt tapasztaltuk, hogy 500-600 Hz alapfrekvencia (a  $h'$  hang) környékén még átlagosan az egyes hangzók zártabbként azonosítása volt jellemző azon ejteni szándékozott magánhangzók esetében, melyek „eredeti” (szándékolt) minőségük szerint nyíltabb és zártabb szomszédokkal is rendelkeznek. Másfelől azonban azt is láttuk, hogy 700 Hz alapfrekvencia ( $f''$ ) magasságától már kifejezetten az egyes énekelt hangzók az ejtési szándéknál nyíltabbként, de különösen  $/n/$  és  $/a/$  hangzókként azonosítása dominált. Ezekről a 700 Hz alapfrekvencia fölött, de különösen a 988 Hz alapfrekvencián (a zenei  $h''$  hang magasságában) tapasztalt azonosításokról azonban már feltehetjük, hogy nem egyszerűen a magánhangzók produkciós nyílását követő perceptuális nyílás eredményei. Az 1000 Hz alapfrekvencia körül ejtett hangzókról ugyanis akusztikai méréseink alapján azt állapítottuk meg, hogy az azok ejtésére jellemző toldalékcső-konfiguráció első és második rezonanciája, azaz az első és második formáns éppen az  $f0$  és a  $2f0$  frekvencia-összetevőket erősítette mind a kilenc magyar magánhangzó esetében. Ez pedig, női ejtésben (a kísérletben részt vevő női beszélők ada-

tai alapján is) az /a:/ beszédbeli megvalósulására jellemző frekvenciaszerkezetet, valamint az /a:/-ra jellemző prominenciaszerkezetet (vö. Traunmüller 1981) jelent (2.5. *alfejezet*).

Megfigyeléseink, valamint az a kísérletünk, hogy Traunmüller (1981) alapvetően a beszédejtésre kidolgozott tonotópiás modelljét alkalmazzuk a magas alapfrekvencián énekelt és  $F1 : f_0$  formánshangolt magánhangzók percepciójára is, arra is következtetni engedtek, hogy az  $F1 : f_0$  hangolás feltehetőleg nem csak a homogén hangszínezet és a nagy hangerejű énekhang produkcióját segíti elő magas alapfrekvenciákon. Vizsgálataink alapján ugyanis azt fogalmazhattuk meg, hogy feltehetőleg szintén az  $F1 : f_0$  hangolás eredménye a középső és alsó nyelvéllés, azaz második és harmadik nyíltsági fokú magánhangzók megjelenése az észleletben 370 Hz alapfrekvencia fölött – a hangolás hiányában ugyanis csak első és negyedik nyíltsági fokú hangzók megjelenését várnánk. Az ily módon második és harmadik nyíltsági fokú magánhangzókként azonosított énekelt hangzók nem feltétlenül (sőt sok esetben nem) az ejtési szándék szerinti azonosítások során jelentek meg válaszként, tehát nem állíthatjuk, hogy az  $F1 : f_0$  hangolás az ejtési szándéknak megfelelő azonosítást támogatta volna. Ugyanakkor talán mégis állítható, hogy az  $F1$ -nek a megemelt, magas alapfrekvenciájú alaphangra hangolásából hasznot hajt az énekes, hiszen ezáltal – legalábbis Traunmüller (1981) vizsgálatai alapján – többfajta magánhangzó-minőség jelenhet meg a percepcióban, mint az  $F1 : f_0$  hangolás nélkül, így az énekprodukción az énekelt hangzók hangszínében változatosabb lehet.

Eredményeink alapján egyértelműen elutasíthatjuk azt a korábban több vizsgálat (Scotto di Carlo–Germain 1985, Benolken–Swanson 1990, Hollien et al. 2000) által is javasolt – és a jelen eredmények fényében igen leegyszerűsítőnek tűnő – javaslatot, hogy az alapfrekvencia emelésével az  $F1 : f_0$  hangolás következtében az egyes ejteni szándékozott, de akusztikailag egyre nyíltabbként, egyre magasabb első formánssal megvalósított magánhangzókat egyre nyíltabb hangzókként is azonosítanák a hallgatók. A jelen vizsgálatok tanúsága szerint ugyanis a produkció és a percepció ennél sokkal komplexebb viszonya mutatkozik meg (az énekelt) magánhangzók azonosításában (is) az alapfrekvencia függvényében.

**H3:** *Mivel magas alapfrekvenciákon az énekelt magánhangzók frekvenciaszerkezete megváltozik, továbbá a magánhangzók spektrális felbontása (a környező mássalhangzókkal alkotott szonoráns*

*hangátmenetek spektrális felbontásával együtt) csökken, a mássalhangzó-környezet a formánsátmenetekben rejlő akusztikai kulcsok révén nem fejt ki pozitív hatást a magánhangzónak az ejtési szándék szerinti azonosítására. Ugyanez érvényes az izolált ejtésű magánhangzók zöngéindítására is. A mássalhangzó-környezet pusztán abban az esetben facilitálhatja az ejtési szándék szerinti azonosítást, ha az azonosítandó magánhangzót tartalmazó hordozó hangsorok olyan értelmes szavak, melyek száma kisebb, mint a válaszként szóba jöhető összes magyar magánhangzó száma.*

Percepció vizsgálataink – a kérdést vizsgáló egyetlen korábbi tanulmány (Smith–Scott 1980) eredményeivel ellentétben – azt bizonyították, hogy sem a mássalhangzó-környezet mint a koartikulációs formánsátmenetek révén dinamikus akusztikai kulcsokat biztosító kontextus, sem pedig az izolált ejtésű magánhangzók zöngéindulása nem fejt ki pozitív hatást a magas alapfrekvencián énekelt magánhangzónak az ejtési szándék szerinti azonosítására (2.3. *alfejezet*). Ugyanakkor azt is kiemeltük, hogy Smith és Scott (1980) valójában – szándékaiktól eltérő módon – nem a dinamikus akusztikai kulcsok, hanem az értelmes szavakba ágyazottság hatását tesztelte és mutatta ki vizsgálatában. Bár a jelen értekezésben vizsgált különböző képzésmódú és némiképp eltérő képzéshelyű nazális és frikatív hangkörnyezetek valamelyest eltérő hatást mutattak a magánhangzó-azonosításra, ezek az eltérések nem voltak számottevőek, így azt állapíthattuk meg, hogy egyik vizsgált mássalhangzó-kontextus sem hatott igazán pozitívan az ejteni szándékozott magánhangzónak az ejtési szándék szerinti azonosítására (2.1. és 2.2. *alfejezet*). Mindez annak fényében nem meglepő, hogy láttuk, az egyes ejteni szándékozott magánhangzók megvalósulása fokozatosan változik, és az ejteni szándékozott magánhangzó-minőségtől egyre inkább eltérő lesz az alapfrekvencia emelésével. Az akusztikus vetületük tekintetében megváltozó hangzók miatt ugyanis ésszerű feltételeznünk, hogy a magánhangzó és a szomszédos mássalhangzók közti formánsátmenetek sem kulcsolhatják az ejteni szándékozott, de meg nem valósított magánhangzót – legfeljebb csak az ejtési szándéktól már többé vagy kevésbé eltérő magánhangzószerű hangot. Ennek értelmében pedig a mássalhangzó-kontextus kizárólag az ejtési szándéktól eltérő azonosítások tendenciáit egységesítheti. Arra, hogy a mássalhangzó-környezet valóban szisztematikusan növelné valamely, az ejtési szándéktól eltérő hangzóazonosítási tendenciát – azaz valamilyen módon diszkretizálná a graduálisan változó hangzóminőségeket – nem találtunk bizonyítékot, így

pusztán a jelen eredmények alapján nem állítható, hogy a mássalhangzó-környezet a formánsátmenetek révén bárminemű pozitív hatással lenne a magánhangzók azonosításának tendenciáira.

Mindemellett azonban azt is láttuk, hogy a mássalhangzó-környezet mint értelmes szavakba ágyazottság nagyban befolyásolta az azonosítás tendenciáit. Bár Smith és Scott (1980) célja nem ez volt, de vizsgálatukkal rávilágítottak, hogy a mássalhangzó-környezet pozitív hatást fejthet ki a magánhangzók az ejtési szándék szerinti azonosítására akkor, ha a magánhangzó és a mássalhangzók együtt értelmes szót alkotnak. A szerzők ugyanis a kontextus hatását egy négy értelmes szót, négy lehetséges kiegészítést, azaz négy magánhangzót mint válaszlehetőséget megengedő feltétel, valamint az angol magánhangzók rendszeréből következően legalább tízféle választ megengedő izolált magánhangzós feltétel összevetésével vizsgálták – és így mutattak ki pozitív hatást az első feltétel javára. A jelen vizsgálatok egyikében éppen ezért gyakorlatilag megismételtük Smith és Scott (1980) vizsgálatát azzal a módosítással, hogy a két feltételhez kiegyenlítettebb (de még mindig nem egyenlő) számú válaszlehetőséget rendeltünk. Az eredmények a várakozásainknak megfelelően azt a feltevés igazolták, hogy megfelelően nagyszámú válaszlehetőség mellett, azaz akkor, ha a hallgatók inkább csak az akusztikus kulcsokra támaszkodhattak az egyes énekelt hangzók azonosításában, a mássalhangzós kontextus elveszítette pozitív hatását, és nem segítette az egyes énekelt hangzók az ejtési szándék szerinti felismerését. A dinamikus akusztikus kulcsok tehát önmagukban nem teszik lehetővé a magánhangzóknak az ejtési szándék szerinti azonosítását (2.4 alfejezet).

Mindezek alapján levontuk azt a következtetést is, hogy szükséges a mássalhangzó-környezet kétértelmű fogalmának pontos definíciója, az egyes tanulmányokban ugyanis sokszor keveredik a két élesen elválasztandó aspektus, a koartikulációs formánsátmenetek és a dinamikus akusztikai kulcsok, valamint az értelmes szavakba ágyazottság hatása. Továbbá vizsgálataink azt is határozottan bizonyították, hogy a magánhangzók akusztikai szerkezetének változása révén – a legtöbb magánhangzó esetében – valóban ellehetetlenül a magánhangzóknak az ejtési szándék szerinti azonosítása a magasabb alapfrekvenciákon. Végül pedig megállapíthattuk, hogy a magas alapfrekvencián énekelt, a hallgató által ismert nyelven megszólaló, értelmes szavakat tartalmazó dalokban és áriákban elhangzó magánhangzók azonosítását elsősorban nem a megvalósuló hangzók akusztikai szerkezete, hanem sokkal inkább csak statisztikai hatások befolyásolják: az egyes, a

magánhangzót tartalmazó szavak előfordulási gyakorisága és valószínűsége, valamint a szavak szomszédsági sűrűsége (*phonological neighborhood density*), azaz egy szó összetéveszthetőségének mértéke, melyet a tőle pusztán a kérdéses magánhangzóban különböző szavak száma határoz meg.

Vizsgálataink demonstrálták, hogy a magánhangzókat az éneklésben jellemző és az alapfrekvencia emelésével egyre fokozódó produkciós változásokat pusztán a beszédpercepciók folyamatok szintje nem képes kompenzálni, a kompenzációhoz a magasabb szintű beszédmegértési folyamatok aktivációja szükséges. Ezt támasztja alá az az eredményünk is, mely szerint az énekesek produkciós gyakorlottságuk ellenére semmilyenfajta percepciók előnyt nem látszanak élvezni a magas alapfrekvencián énekelt magánhangzóknak az ejtési szándék szerinti azonosításában.

### **3.2. A vizsgálatok közvetlen hozadéka, kitekintés**

Bár a valóságban az alapszófajú, több szótagos szavak szomszédsági sűrűsége (*phonological neighborhood density*) alacsonyabb, mint a jelen kísérletben vizsgált szavaké, valamint alapvetően nem jellemző az értelmetlen szavak ejtése a klasszikus operaéneklés során, az is igaz, hogy a szavakat elég erősen szótagolva ejtik az énekesek, ráadásul úgy, hogy a szótagokból több is a kritikus alapfrekvenciákon felül valósul meg. Így az értelmetlen szótagokban vagy izolációban ejtett magánhangzók azonosítását vizsgáló kísérleteknek nem csak azért van relevanciája, mert az akusztikai elemzésen alapuló észlelés, a beszédészlelés a beszédmegértés alapját jelenti, hanem azért is, mert valószínűleg nem ritka, hogy a valóságban is mesterségesen előállított körülményekhez nagyon hasonló körülmények között kell azonosítanunk egy-egy szótagot az éneklés hallgatása során. Mindezek miatt – bár első közelítésben furcsának hathat –, az itt tárgyaltak függvényében feltehetőleg jó és támogatandó gyakorlatnak tekinthető az operaszövegek feliratozása még a beszélők anyanyelvén megszólaló előadások esetén is, valamint a dalesteken, koncerteken a hallgatóság számára valamilyen formában prezentált szöveggönyv. Ezen túlmenően pedig nem szabad megfeledkeznünk arról sem, hogy a magánhangzók alapfrekvenciafüggő megvalósítása szem előtt tartandó szempont lehet(ne) az operaszövegek fordításában is. Míg egyes kutató-

sok szerint például Wagner kifejezetten ügyelt rá, hogy a magasabb alaphfrekvenciákon harmadik és negyedik nyíltsági fokú (jellemzően /a/) hangzókkal „könnyítse meg az énekes dolgát” (Smith–Wolfe 2009), egy rosszul sikerült fordítás könnyedén semmissé teheti ezt a törekvést.

Elgondolkodásra érdemes kérdés, hogy eredményeink mennyiben vehetők figyelembe az énektanítás módszertanában. Tudjuk, hogy sokszor bizonyos alapvetően tudatosan „nehezen elérhető” artikulációs mozgások – például a gége függőleges helyzetének direkt változtatása – csak akkor hajthatók végre hatékonyan, ha az ejtéshez kiadott tudatos produkciós parancsok nem közvetlenül az artikulációs szervek működését igyekeznek befolyásolni. (Ezt használják ki például a fül-orr-gégészek mindennapos praxisukban, mikor a gégetükrözéshez az /i/ hangzó ejtésére kérik a páciens-t úgy, hogy közben leszorítják a nyelvet, és szétfeszítik az ajakat, tehát teljesen ellehetetlenítik az ejtést. Így érik el ugyanis, hogy a páciens a lehető legmagasabb gégehelyzetben fonáljon, ami a tükör számára is látható magasságba emeli a gégét – míg a gége emelésére valószínűleg hiába kérnék a beteget.) A különböző, az énekhang minősége szempontjából fontos produkciós jellemzők, akusztikus hatások eléréséhez tehát feltehetően szükséges az az illúzió, mely szerint az egyes magánhangzók még magas alaphfrekvenciákon is elkülöníthetőek a produkcióban. Ugyanakkor az is bizonyos, hogy az énekeseket nagymértékben megnyugtatja egy, a hangzók elkülönítésének lehetetlenségéről szóló tudományos bizonyíték is. Jól példázza ezt egy énekes (nem akusztikai-fonetikai irányultságú) disszertációjának pár sora: „a tanult technikák segítségével biztonsággal létrehozhatjuk azt a hangot, ami megfelelő dinamikával hallható a nagyzenekari hangzás mellett, de rendkívül magas hangtartományban és a hosszan tartott hangzás esetén könnyen szövegtorzulást eredményez. Ez nemcsak a zene és a prozódia (esetleges fordítás) szerkesztési problémája, hanem teljesen normális akusztikai folyamat eredménye” (Váradi 2010: 36). Ráadásul az énekes szakma berkeiben is gyakran megjelenő kritika az énektechnika tanításával kapcsolatban, hogy az nehezen lefordítható, és nem mindig és mindenki számára jól megérthető absztraktabb metaforákat használ az egyes produkciós, akusztikus jellegek kialakítására (pl. „elől képzett” hang, „nyitott” hang stb.). Ez ugyanis az eltérő fiziológiájú énekesek esetében komoly problémát okozhat – énekesek személyes beszámolóí szerint nem ritka, hogy egy énektanár és egy diák közti munkakapcsolat a kettejük közti testi, fiziológiai különbségek miatt hiúsul meg. Ráadásul a beszédképző szervek anatómiájára és fiziológiájára támaszkodó oktatásnak, tehát annak a módszertannak,

mely az előbb említett különbségekből fakadó problémákat a fizikai valóságot megragadó instrukciók révén áthidalni képes, már van hagyománya. Nagyban ezen alapszik ugyanis az iskolateremtő Jo Estill vagy Cornelius Reid oktatásmódszertana is.

### 3.3. Tézisek

1. Igazoltuk azt a magas alapfrekvenciákon éneklő szoprán énekesekkel kapcsolatban feltételezett stratégiát a magyar magánhangzókra, mely szerint akkor, ha az alapfrekvencia emelése során az alapfrekvencia magasabb lenne, mint az első formáns, az énekes az első formáns frekvenciáját az alapfrekvencia értékére emeli. Demonstráltuk az összes magyar magánhangzó produkciós nyitódását az alapfrekvencia emelésével, valamint a magánhangzók redukcióját és konvergálását az /a:/ pozíciója felé a magánhangzótérben.

2. Sikeresen teszteltük a manuális inverz szűrés módszerét elektrolottográffal, valamint elektromos gégével kiegészítve a magas alapfrekvenciájú szopránénekülésben realizálódó magánhangzók formánsainak a fonáció alapfrekvenciájától független mérésében.

3. Megcáfoltuk azt a szakirodalomban honoló (de nem jól kontrollált kísérletekre alapozott) feltételezést, mely szerint a magánhangzók az alapfrekvencia emelésével fokozódó produkciós nyitódását a hangzómegvalósulások fokozatos percepciós nyitódása kísérné. Eredményeink szerint a produkció és percepció összefüggése e kérdésben (is) jóval komplexebb. Az alapfrekvencia emelésével alapvetően a magánhangzók percepciós minőségének elbizonytalanodását tapasztaltuk.

3.a A produkciós adatoknak megfelelő, de az ejtési szándéknál nyíltabb (/p a:/ hangzóként történt) magánhangzó-azonosításokat csak abban az esetben találtunk, amikor az alapfrekvencia kellően magas volt ahhoz, hogy az  $f_0$  és  $2f_0$  egymáshoz viszonyított távolsága, valamint a két összetevő erősítettsége is hasonlóvá tegye az egyes magánhangzószerű hangzókat az /a:/ beszédbeni prominenciamintázatához. Ez az  $F_1 : f_0$  hangolás miatt csak akkor valósul meg egyértelműen, amikor az alapfrekvencia emelésével változó frekvenciaértékű  $F_2$  formáns eléri (és erősíteni kezdi) a  $2f_0$  összetevőt.

3.b Az 1000 Hz alatti alapfrekvenciákon a produkció és percepció az egyszerű lineáris összefüggésnél sokkal komplexebbnek mutatkozik: 5-600 Hz-ig még alapvetően az egyes „eredeti” minőségük szerint nyíltabbként és zárttabbként is azonosítható hangzók az ejtési szándéknál zár-



tabbként azonosítása dominál, az ejtési szándéknál nyíltabb azonosítások száma csak e fölött a frekvenciasáv fölött kezd el növekedni. Feltehetjük, hogy a nyíltabbként azonosításhoz e produkciósan nyíltabb és  $F1 : f_0$  formánshangolt hangzók esetében szükséges a spektrális alul-mintavételezettség fokozódása, és az  $F2$  és  $2f_0$  egybeesése (azaz a  $2f_0$  erősítése), tehát a percepció nyitódás nem a produkció nyitódás egyenes következménye.

4. Valószínűnek látszik, hogy éppen a magánhangzók akusztikai szerkezetét megváltoztató  $F1 : f_0$  hangolás következtében lehetséges a második és harmadik nyíltsági fokú magánhangzók megjelenése a percepcióban (bár nem feltétlenül az ejtési szándéknak megfelelően) – a formánshangolási tendencia nélkül elsősorban első és negyedik nyíltsági fokú hangzók megjelenését jósolja a szakirodalom.

5. Megállapítottuk, hogy a mássalhangzó-környezet a dinamikus akusztikus kulcsok (azaz a koartikulációs formánsátmenetek révén) nem segíti a magas alapfrekvencián énekelt magánhangzók az ejtési szándék szerinti azonosítását – szemben a beszéddel. Ugyanakkor a mássalhangzó-környezet mint értelmes szavakba ágyazottság pozitív hatást fejthet ki akkor, ha az adott szó valószínűsége magas, illetve a szónak kevés fonológiai szomszédja van (tehát csak kis számú olyan szószomszédal rendelkezik, melytől csak a kérdéses magánhangzó különbözteti meg).

6. Megállapítottuk, hogy az énekes és nem énekes (tehát énektechnikailag képzetlen) hallgatóság gyakorlatilag ugyanolyan tendenciák szerint azonosítja a magas alapfrekvencián énekelt magánhangzókat, és az éneklésben szerzett gyakorlat nem jelent percepció előnyt a magánhangzók az ejtési szándék szerinti felismerésében.

7. Megállapítható, hogy az énekelt magánhangzók akusztikai szerkezete az alapfrekvencia emelésével egyre jobban megváltozik (eltér a beszédbeni megvalósulásától), mely produkciós tendencia pusztán a beszédpercepció szintjén nem kompenzálható – a kompenzációhoz a magasabb szintű beszédmegértési folyamatok aktivációja szükséges.

## Irodalom

- Ashby, Patricia 2011. *Understanding Phonetics*. Hodder Education, London.
- American Standards Association 1960. American Standard Acoustical Terminology. New York: American Standards Association.
- Beddor, Patrice S. 1983. *Phonological and Phonetic Effects of Nasalization on Vowel Height*. PhD értekezés, Indiana University Linguistics Club, Bloomington.
- von Békésy, Georg 1960. *Experiments in hearing*. McGraw-Hill, New York.
- Benade, A. H. 1976. Fundamentals of musical acoustics, Oxford University Press, London.
- Benolken, Martha S. – Swanson, Charles E. 1990. The effect of pitch-related changes on the perception of sung vowels. *Journal of the Acoustical Society of America* 87(4). 1781–1785.
- Berger, Kenneth W. 1964. Some factors in the recognition of timbre. *Journal of the Acoustical Society of America* 26(10). 1888–1891.
- Boersma, Paul – Weenink, David 2009. *Praat: Doing phonetics by computer (Version 5.3)*. <http://www.praat.org>. (A letöltés ideje 2014. január 12.)
- Branderud, Peter – McAllister, Robert – Kassling, Bo 1994. Methodological studies of Movetrack: coil tilt and placement. Stockholm University: PERILUS XVIII. 91–109.
- Bresch, Erik –Narayanan, Shrikanth 2010. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *Journal of the Acoustical Society of America* 128(5). EL335–EL341.
- Bóna Judit – Imre Angéla 2010. A rövid–hosszú magánhangzók óvodás és kisiskolás gyermekek beszédprodukciónak. In: Navracsics Judit (szerk.) *Nyelv, beszéd, írás: Pszicholingvisztikai tanulmányok I.* Budapest, Tinta Könyvkiadó. 49–56.
- Bóna Judit 2012. A zöngétlen explozívák zöngékezdési ideje idősek és fiatalok felolvasásában. In: Gósy Mária (szerk.) *Beszéd, adatbázis, kutatások*. Budapest, Akadémiai Kiadó. 121–137.
- Bóna Judit 2014. Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America* 136(2). EL116–EL121.

- Carlson, Rolf – Granström, Björn – Fant, Gunnar 1970. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 2–3. 19–35.
- Carlson, Rolf – Fant, Gunnar – Granström, Björn 1975. Two-formant models, pitch and vowel perception. In: Fant, Gunnar – Tatham, M. A. A. (szerk.) *Auditory Analysis and Perception of Speech*, London: Academic Press. 55–82.
- Carlsson-Berndtsson, Gunilla – Sundberg, Johan 1991. Formant frequency tuning in singing. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 32(1). 29–35.
- Chiba, Tsutomu – Kajiyama, Masato 1941. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan, Tokyo, Japan.
- Chistovich, Ludmilla A. – Lublinskaya, Valentina V. 1979 The ‘center of gravity’ effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research* 1(3). 185–195.
- Cooker, Harry S. 1973 On the problem of tracking mandibular movements. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 14(2–3). 001–012.
- de Cheveigné, Alain – Kawahara, Hideki 1999. Missing-data model of vowel identification. *Journal of the Acoustical Society of America* 105(6). 3497–3508.
- Deme Andrea – Grácsi Tekla Etelka – Jankovics Júlia 2013. Obstruent voicing in singing. (előadás) *15th Summer School of Psycholinguistics; 2013 May 26–30*, Balatonalmádi.
- Echternach, Matthias – Sundberg, Johan – Arndt, Susan – Markl, Michael – Schumacher, Martin – Richter, Bernhard. 2010. Vocal tract in female registers – A dynamic real-time MRI study. *Journal of Voice* 24(6). 133–139.
- Epps, J. – Smith, John R. – Wolfe, Joe 1997. A novel instrument to measure acoustic resonances of the vocal tract during speech. *Measurement Science and Technology* 8. 1112–1121.
- Evgrafova, Karina – Evdokimova, Vera 2012. Perception of russian vowels in singing. *Baltic HLT Frontiers in Artificial Intelligence and Applications*. IOS Press 247. 42–9.
- Fant, Gunnar 1959. *Acoustic Analysis and Synthesis of Speech with Applications to Swedish*. Ericsson Technics No. 1.

- Fant, Gunnar 1960. *Acoustic Theory of Speech Production*. Mouton & Co., The Hague, Netherlands.
- Flemming, Edward. 24.963 Linguistic Phonetics, Fall 2005. (Massachusetts Institute of Technology: MIT OpenCourseWare), <http://ocw.mit.edu> (Accessed 15 March, 2015). License: Creative Commons BY-NC-SA.
- Garcia, Manuel (1911). *Garcia's Treatise on the Art of Singing*. London: Leonard & Co.
- Gósy Mária 1989 *Beszédészlelés*. Linguistica Series A, Studia et Dissertationes 2. MTA Nyelvtudományi Intézet, Budapest.
- Gósy Mária. 1999. Az egyéni hangszínezet és a beszélő felismerésének kísérleti-fonetikai megközelítése. *Magyar Nyelvőr* 123. 424–438.  
<http://www.c3.hu/~nyelvor/period/1234/123406.htm>. (A letöltés ideje: 2015. február 05.)
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária 2012. Az alsóbb nyelvtudomány magyar magánhangzók formánsszerkezete. In: Gósy Mária (szerk.) *Beszéd, adatbázis, kutatások*. Budapest, Akadémiai Kiadó. 43–66.
- Gósy, Mária 2013. Inter-speaker and intra-speaker variability indicating a synchronous speech sound change. In: Szigetvári Péter (szerk.) *VLLxx: Papers in linguistics presented to László Varga on his 70th birthday*. Budapest, Tinta Könyvkiadó. 313–332.
- Gósy Mária – Bóna Judit 2014. Magánhangzók ejtése fiatalok és idősek spontán beszédében. *Magyar Nyelv* 110(2). 129–143.
- Garnier, Maëva – Henrich, Natalie – Smith, John – Wolfe, Joe 2010. Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America* 127(6). 3771–3780.
- Gottfried, Terry L. – Chew, Stephen L. 1986. Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America* 79(1). 124–130.
- Grácsi Tekla Etelka – Hovráth Viktóra 2010. A magánhangzók realizációja spontán beszédben. *Beszédkutatás 2010*. 5–16.
- Gramming, Patricia – Nord, Lennart – Sundberg, Johan – Eliot, Ninni 1993. Does the nose resonate during singing? *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 34(4). 35–42.

- Henrich, Nathalie – d'Alessandro, Christophe – Doval, Boris – Castellengo, Michèle 2004. On use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America* 115(3). 1321–1332.
- Hertegård, S. and J. Gauffin 1993. Voice source – vocal tract interaction during high-pitched female singing. In: *Proceedings of the Stockholm Music Acoustics Conference 1993 (SMAC 1993)*. Stockholm, Royal Swedish Academy of Music. 177–182.
- Herbst, Christian T. – Fitch, W. Tecumseh S. –Svec, Jan G. 2010. Electroglottographic wavegrams: a technique for visualizing vocal fold dynamics noninvasively. *The Journal of the Acoustical Society of America* 128(5). 3070–3078.
- Hickok, Gregory. 2010. The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes* 25(6). 1–28
- Hollien, Harry – Mendes-Scwartz, Ana P. – Nielsen, Kenneth 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14(2). 287–298.
- Honda Kiyoshi 1983. Relationship between pitch control and vowel articulation. In: Studdert-Kennedy, Michael - O'Brian, Nancy (szerk.) *Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications*. Haskins Laboratories, New Haven. 269–282.
- Honda, Kyioshi é. n. Control of Hyoid-Larynx Complex by Jo Estill (előadás).
- Kovács Magdolna 1998. A spektrális minőség és az időtartam szerepe a magánhangzók percepciójában. *Beszéd kutatás 1998*. MTA Nyelvtudományi Intézet, Budapest. 35–46.
- Hoole, Philip – Kroos, Christian 1998. Control of larynx height in vowel production. In: *Proceedings of the 5th Conference on Language Processing (ICSLP) 2*. 531–534.
- Horváth Viktória 2005. A magánhangzók nazalizációjáról. In: Gósy Mária (szerk.) *Beszéd kutatás 2005*. MTA Nyelvtudományi Intézet, Budapest. 51–62.
- Horváth Viktória 2008. Nazális hatás a magánhangzóképzés függvényében. In: Gósy Mária (szerk.) *Beszéd kutatás 2008*. MTA Nyelvtudományi Intézet, Budapest. 22–32.
- Hurme, Pertti – Sonninen, Aato 1995. Vertical and saggital position of the larynx in singing. In: Elenius, K. and Branderud, P. (szerk.) *Proceedings of the XIII International Congress of Phonetic Sciences*. 214–217.

- International Phonetic Association 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, Cambridge University, Cambridge.
- International Phonetic Association 2005. *The International Phonetic Alphabet* (revised to 2005). [https://www.internationalphoneticassociation.org/sites/default/files/IPA\\_chart\\_\(C\)2005.pdf](https://www.internationalphoneticassociation.org/sites/default/files/IPA_chart_(C)2005.pdf). (A letöltés ideje: 2015. február 15.)
- Joliveau, Elodie – Smith, John – Wolfe, Joe 2004. Vocal tract resonances in singing: the soprano voice. *Journal of the Acoustical Society of America* 116(4). 2434–2439.
- Jenkins, James J. – Strange, Winifred – Edman, Thomas R. 1983. Identification of vowels in "vowelless" syllables. *Perception & Psychophysics* 34(5). 441–450.
- Johansson, Christine – Sundberg, Johan – Wilbrand, Hermann 1982. X-ray study of articulation and formant frequencies in two female singers. In: Askenfell, Anders – Felicetti, S. – Jansson, Erik – Sundberg, Johan (szerk.) *Proceedings of the Stockholm Music Acoustics Conference 1983 (SMAC 1983)*, Stockholm, Royal Swedish Academy of Music. 203–218.
- Kállai János – Bende István – Karádi Kázmér – Racsmány Mihály 2008. *Bevezetés a neuropszichológiába*. Medicina, Budapest.
- Kerényi Miklós György 1959. *Az éneklés művészete és pedagógiája*. Zeneműkiadó, Budapest.
- Kovács Magdolna 1998. A spektrális minőség és az időtartam szerepe a magánhangzók percepciójában. *Beszéd kutatás*, MTA Nyelvtudományi Intézet, Budapest. 35–46.
- Krakow, Rena A. – Beddor, Patrice S. – Goldstein, Louis M. – Fowler, Carol A. 1987. Coarticulatory influences on the perceived height of nasal vowels. *Status Report on Speech Research*, Haskins Laboratories. 31–53.
- Ladefoged, Peter – Broadbent, Donald E. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America* 29(1). 98–104.
- Ladefoged, Peter 1967. *Three areas of Experimental phonetics*. Oxford University Press, London. 62–75.
- Lee, Sungbok – Potamianos, Alexandros – Narayanan, Shrikanth 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105(3). 1455–1468.

- Lieberman, Alvin M. – Mattingly, Ignatius G. 1985. The motor theory of speech perception revised. *Cognition* 21(1). 1–36.
- Lindblom, Björn 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, William – Marchal, Alain (szerk). *Speech Production and Speech Modeling*. Kluwer, Dordrecht. 403–439.
- Lindblom, Björn E. – Sundberg, Johan 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America* 50(4). 1166–1179.
- Lindblom, Björn – Brownlee, S. – Davis, Barbara – Moon, Seung-Jae 1992. Speech transforms. In *Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication 1991 (PPoSpSt-1991)*, Barcelona, Catalonia, Spain. 357–368. [http://www.isca-speech.org/archive\\_open/ppospst](http://www.isca-speech.org/archive_open/ppospst)
- Mády Katalin 2008a. Beszédpercepció és pszicholingvisztika. [http://www.phonetik.uni-muenchen.de/~mady/pub/mady\\_percepcio.pdf](http://www.phonetik.uni-muenchen.de/~mady/pub/mady_percepcio.pdf). (A letöltés ideje: 2015. február 15.)
- Mády Katalin 2008b. Magyar magánhangzók vizsgálata elektromágneses artikulográffal gyors és lassú beszédben. In: Gósy Mária (szerk.) *Beszéd kutatás 2008*. MTA Nyelvtudományi Intézet, Budapest. 52–66.
- Markó, Alexandra – Grácz, Tekla Etelka – Bóna, Judit 2010. The realization of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57(2-3). 210–238.
- Markó Alexandra 2013. *Az irreguláris zöngé funkciói a magyar beszédben*. ELTE Eötvös Kiadó, Budapest.
- McKinney, James 1994. *The Diagnosis and Correction of Vocal Faults*. Genovex Music Group.
- Mitchell, Helen 2005. *Defining vocal quality in female classical singers: pedagogical, acoustical and perceptual studies*. Doktori értekezés, University of Sydney, Sydney.
- Nádasdy Ádám – Siptár Péter 1994. A magánhangzók. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 42–182.
- Németh Géza – Olasz Gábor (szerk.) 2010. *A magyar beszéd. Beszéd kutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest.

- Vicsi Klára 2010a. A hallási folyamat. In: Németh Géza – Olasz Gábor (szerk.) 2010. *A magyar beszéd. Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest. 27–38.
- Vicsi Klára 2010b. A beszéd fizikai jellemzése. In: Németh Géza – Olasz Gábor (szerk.) 2010. *A magyar beszéd. Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest. 38–56.
- Neary, Terence M. 1989. Static, dynamic, and relational properties in vowel perception, *Journal of the Acoustical Society of America* 85(5). 2088–2113.
- Peterson, Gordon E. – Barney, Harold L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24(2). 175–184.
- Peterson, Gordon E. 1961. Parameters of vowel quality. *Journal of Speech and Hearing Research* 4(1). 10–29.
- Pabst, Friedemann – Sundberg, Johan 1992. Tracking multi-channel electroglottograph measurement of larynx height in singers. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 33(2-3). 67–78.
- R Core Team 2013. *R: A Language and Environment for Statistical Computing*, Vienna, Austria. <http://www.R-project.org>
- Rogers, Catherine, L. – DeMasi, Teresa, M. – Krause, Jean, C. 2010. Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustical Society of America* 128(1). 410–423.
- Rosner, B. S. – Pickering, J. B. 1994. *Vowel perception and production*. Oxford University Press, Oxford.
- Roubeau, Bernard – Henrich, Nathalie – Castellengo, Michèle 2009. Laryngeal vibratory mechanisms: The notion of vocal register revisited. *Journal of Voice* 23(4). 425–438.
- Ryalls, John H. – Libermann, Philip 1983. Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America* 72(5). 1631–1634.
- Saldanha, E. L. – Corso, John F. 1964. Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America* 36(11). 2021–2026.



- Scotto di Carlo, Nicole – Germain, Aline 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 42(2).188–97.
- Scotto di Carlo, Nicole 2007. Effect of multifactorial constraints on intelligibility of opera singing (II). *Journal of Singing* 63(4). 559–567.
- Seikel, Anthony J. – King, Douglas, W. – Drumright, David G. 2010. *Anatomy & Physiology for Speech, Language and Hearing* (4th ed), Clifton Park, NY, Delmar Cengage Learning.
- Siptár, Péter – Törkenczy, Miklós 2000. *The Phonology of Hungarian*. Oxford University Press, Oxford.
- Siptár Péter 2014. Fonológiai ábrázolás és fonetikai megvalósítás: a magyar /a:/ esete. *Magyar Nyelv* 110(3). 335–342.
- Sjölander, Kåre – Beskov, Jonas 2000. Wavesurfer – an open source speech tool. In: Yuan, B. – Huang, T. – Tang, X. (szerk.) *Proceedings of ICSLP 2000, 6th International Conference on Spoken Language Processing*, Beijing. 464–467. <http://www.speech.kth.se/wavesurfer> (A letöltés ideje: 2013. február 14.)
- Smith, Lloyd – Scott, Brian L 1980. Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America* 67(5). 1795–1797.
- Smith, John – Wolfe, Joe 2009. Vowel-pitch matching in Wagner’s operas: Implications for the intelligibility and ease of singing. *Journal of the Acoustical Society of America* 125(5). 196–201.
- Sundberg, Johan 1969. Articulatory differences between spoken and sung vowels in singers. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 10(1). 33–46.
- Sundberg, Johan 1970. Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica* 22(1). 28–48.
- Sundberg, Johan 1972. An articulatory interpretation of the ‘singing formant’. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 13(1). 45–53.
- Sundberg, Johan 1975. Formant technique in a professional female singer. *Acta Acustica united with Acustica* 32(2). 89–96.
- Sundberg, Johan 1977. The acoustics of the singing voice. *Scientific American* 234(3). 82–91.

- Sundberg, Johan 1979. Perception of singing. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 20(1). 1–48.
- Sundberg, Johan 1987. *The science of the singing voice*. Northern Illinois University Press, DeKalb Illinois.
- Sundberg, Johan 1994. Perceptual aspects of singing. *Journal of Voice* 8(2). 106–122.
- Sundberg, Johan 2001. Level and center frequency of the singer's formant. *Journal of Voice* 15(2). 176–186.
- Sundberg, Johan 2009. Articulatory configuration and pitch in a classically trained soprano singer. *Journal of Voice* 23(5). 546–551.
- Sundberg, Johan – Nordström, P-E. 1976. Raised and lowered larynx – the effect on vowel formant frequencies. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 17(2–3). 35–39.
- Sundberg, Johan – Askenfelt, Anders 1981. Larynx height and voice source. A relationship? *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 22(2–3). 23–36.
- Sundberg, Johan – Leanderson, R. – Euler, Curt von 1988. Activity relationship between diaphragm and cricothyroid muscles. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 29(2–3). 83–91.
- Sundberg, Johan – Lã, Filipa – Gill, Brian P. 2013. Formant tuning strategies in professional male opera singers. *Journal of Voice* 27(3). 278–288.
- Sundberg, Johan –Skoog, Jörgen 1995. Jaw opening, vowel and pitch. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 36(2–3). 43–50.
- Srulovicz, Peter – Goldstein, Julius L. 1983. A central spectrum model: A synthesis of auditory-nerve timing and place cues in nonaural communication of frequency spectrum. *Journal of the Acoustical Society of America* 73(4). 1266–1276.
- Standards Secretariat, Acoustical Society of America, 1994. ANSI S1.1-1994 (R2004) American National Standard Acoustical Terminology, (12.41) Acoustical Society of America, Melville, NY.

- Stephen, F. A. 1996. Movement of velum during speech and singing in classically trained singers. *Journal of Voice* 11(2). 212–221.
- Stevens, Kenneth Noble – House, Arthur S. 1961. An acoustical theory of vowel production and some of its implications. *Journal of Speech, Language and Hearing Research* 4(4). 303–320.
- Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3–45
- Stevens, Kenneth N. 1998. *Acoustic Phonetics*. The MIT Press, Cambridge, Massachusetts–London.
- Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111(4). 1872–1891.
- Strange, Winifred – Verbrugge, Robert R. – Shankweiler, Donald P. – Edman, Thomas R. 1976. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America* 60(1). 213–224.
- Svante Granqvist. é. n. *DeCap*. tolván.com. (A letöltés ideje: 2013. április 26.)
- Syrdal, Ann K. – Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086–1100.
- Szaszák György 2010. A beszéd számítógépes feldolgozása. In: Németh Géza – Olasz Gábor (szerk.) 2010. *A magyar beszéd. Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest. 209–243.
- Szende, Tamás 1994. Illustrations of the IPA: Hungarian. *Journal of the International Phonetic Alphabet* 24(2). 91–94.
- Tajima, Keiichi – Tanaka, Kuniyoshi – Martin, Andrew – Mazuka, Reiko 2013. Is the vowel length contrast in Japanese exaggerated in infant-directed speech? *Proceedings of InterSpeech 2013*. 3211–3215
- Tarnóczy Tamás 1982. *Zenei akusztika*. Zeneműkiadó, Budapest.
- Thorpe, William C. – Cala, Stephen J. – Chapman, Janice – Davis, Pamela J. 2001. Patterns of breath support in projection of the singing voice. *Journal of Voice* 15(1). 86–104.
- Titze, Ingo R. 1988. The physics of small amplitude oscillations of the vocal folds. *Journal of the Acoustical Society of America* 83(4). 1536–1552.

- Titze, Ingo. R. 1994. *Principles of Voice Production*. Prentice-Hall, Englewood Cliffs, NJ.
- Titze, Ingo R. 1998. Voice research: the wide pharynx. *Journal of Singing* 55(1), 27–28.
- Titze, Ingo R. 2001. Acoustic interpretation of resonant voice. *Journal of Voice* 15(4). 519–528.
- Titze, Ingo R. 2004. A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice. *Journal of Voice* 18(3). 292–298.
- Titze, I. R. 2008. Nonlinear source-filter coupling in phonation: Theory. *Journal of the Acoustical Society of America* 123(5). 2733–2749.
- Titze, Ingo, R. – Story, Brad H. 1997. Acoustic interactions of the voice source with the lower vocal tract. *Journal of the Acoustical Society of America* 101(4). 2234–2243.
- Trautmüller, Hartmut 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69(5). 1465–1475.
- Váradí Marianna 2010. *A művészi hangképzés kérdései, és néhány példa a magyar operafordítások énekelhetőségének problematikájáról*. Doktori értekezés, Liszt Ferenc Zeneművészeti Egyetem, Budapest.
- Vennard, William 1964. An experiment to evaluate the importance of nasal resonance in singing. *Folia Phoniatrica* 16. 146–153.
- Vennard, William 1967. *Singing, the mechanism and the technique*. Carl Fisher, New York.
- Vorperian, Hourí K. – Kent, Ray D. 2007. Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research* 50. 1510–1545.
- Wassink, Alicia Beckford – Wright, Richard A. – Franklin, Amber D. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35(3). 363–379.
- Weiss, Rudolf – Brown, W. S., Jr. – Morris, Jack 2001. Singer's formant in sopranos: fact or fiction? *Journal of Voice* 15(4). 457–468.
- Wendahl, Ronald R. 1959. Fundamental frequency and absolute vowel identification. *Journal of Acoustical Society of America* 31(1). 109–110.
- Wodarcz-Magdics, Klára 1970. Experiments in Hungarian vowel recognition. *Linguistics* 56. 64–87.

- Wolfe, Joe – Garnier, Maëva – Smith, John 2009. Vocal tract resonances in speech, singing and playing musical instruments. *Human Frontier and Science Journal* 3. 6–23.
- Wood, Sidney 1975. The weakness of the tongue-arching model of vowel articulation. *Lund Working Papers* 11. 55–107.
- Wright, James 1975. Effects of vowel nasalization on the perception of vowel height. In: Ferguson, C.A. – Hyman, L. M. – Ohala, John J. (szerk.) *Nasalfest: Papers from a Symposium on Nasals and Nasalization (Language Universals Project)*, Stanford University, Stanford. 373–388.
- Yanagisawa, Eiji – Kmucha Steven T. – Estill, Jo 1990. Role of the soft palate in laryngeal functions and selected voice qualities. *Annals of Otology, Rhinology, and Laryngology* 99(1). 18–28.
- Yanagisawa, Eiji – Estill, Jo – Mambrino, Lawrence – Talkin, David 1991. Supraglottic contributions to pitch raising. Videoscopic study with spectroanalysis. *Annals of Otology, Rhinology, and Laryngology* 100(1). 19–30.
- Zenker, Wolfgang – Zenker, Adolf 1960. Über die Regelung der Stimmröhrenspannung durch von aussen eingreifende Mechanismen. *Folia Phoniatrica* 12(1). 1–36.
- Zwicker, E. – Terhardt, E. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of Acoustical Society of America* 68(5). 1523–1525.

## A szerző az értekezésben felhasznált publikációinak listája

- Deme Andrea 2011. Az énekelt magánhangzók fonetikai elemzése. In: Parapatics Andrea (szerk.) *Félúton 6: A hatodik Félúton konferencia (2010) kiadványa*. 25-42.
- Deme Andrea 2011. Az énekelt magánhangzók észlelése réshangkörnyezetben. In: Váradi T (szerk.) *V. Alkalmazott Nyelvészeti Doktorandusz Konferencia*. MTA Nyelvtudományi Intézet, Budapest. 16-28.
- Deme Andrea 2012. Az énekelt magánhangzók fonetikai elemzése. In: Parapatics Andrea (szerk.) *Doktoranduszok a nyelvtudomány útjain: 6. Félúton konferencia*. ELTE – Eötvös Kiadó (Talentum), Budapest. 33-46.
- Deme, Andrea 2013. On the Hungarian sung vowels. *The Phonetician* 105-106(1–2). 73-87.
- Deme Andrea 2014a. A mássalhangzó-környezet és a zöngéindulás hatása az énekelt magánhangzók észlelésére. *Beszéd kutatás* 22. MTA Nyelvtudományi Intézet, Budapest. 155-171.
- Deme, Andrea 2014b. Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing. *Journal of Voice* 28(4). 523.e19-523.e25.
- Deme, Andrea 2014c. Formant strategies of professional female singers at high fundamental frequencies. In: Fuchs, Susanne –Grice, Martine –Hermes, Anne –Lancia, Leonardo –Mücke, Doris (szerk) *Proceedings of the 10th International Seminar on Speech Production Cologne (ISSP)*, Köln. 90–93.
- Deme Andrea (megjelenőben). Speech percpetion at its best: Extracting linguistic information from phonetically underspecified imput. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS) 2015*.